Perturbative formulation of general continuous-time Markov model of sequence evolution via insertions/deletions, Part II: Perturbation analyses

Kiyoshi Ezawa [1, 2, *], Dan Graur [1], and Giddy Landan [1, 3]

[1] Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001, USA
[2] Present address: Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka 820-8502, JAPAN
[3] Present address: Institute of Genomic Microbiology, Heinrich-Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, GERMANY

[*] Corresponding Author
(E-mail: kezawa@bio.kyutech.ac.jp, kezawa.ezawa3@gmail.com)

Abstract

Background

Insertions and deletions (indels) account for more nucleotide differences between two related DNA sequences than substitutions do, and thus it is imperative to develop a stochastic evolutionary model that enables us to reliably calculate the probability of the sequence evolution through indel processes. In a separate paper (Ezawa, Graur and Landan 2015a), we established a theoretical basis of our *ab initio* perturbative formulation of a *genuine* evolutionary model, more specifically, a continuous-time Markov model of the evolution of an *entire* sequence via insertions and deletions. And we showed that, under some conditions, the *ab initio* probability of an alignment can be factorized into the product of an overall factor and contributions from regions (or local alignments) separated by gapless columns.

Results

This paper describes how our *ab initio* perturbative formulation can be concretely used to approximately calculate the probabilities of all types of local pairwise alignments (PWAs) and some typical types of local multiple sequence alignments (MSAs). For each local alignment type, we calculated the fewest-indel contribution and the next-fewest-indel contribution to its probability, and we compared them under various conditions. We also derived a system of integral equations that can be numerically solved to give "exact solutions" for some common types of local PWAs. And we compared the obtained "exact solutions" with the fewest-indel contributions. The results indicated that even the fewest-indel terms alone can quite accurately approximate the probabilities of local alignments, as long as the segments and the branches in the tree are of modest lengths. Moreover, in the light of our formulation, we examined parameter regions where other indel models can safely approximate the correct evolutionary probabilities. The analyses also suggested some modifications necessary for these models to improve the accuracy of their probability estimations.

Conclusions

At least under modest conditions, our *ab initio* perturbative formulation can quite accurately calculate alignment probabilities under biologically realistic indel models. It also provides a sound reference point that other indel models can be compared to. [This paper and three other papers (Ezawa, Graur and Landan 2015a,b,c) describe a series of our efforts to develop, apply, and extend the *ab initio* perturbative formulation of a general continuous-time Markov model of indels.]

Keywords

Insertions/deletions (indels), pairwise sequence alignment (PWA), multiple sequence alignment (MSA), probability, likelihood, continuous-time Markov model, perturbation theory, power-law length distribution, exact solution

List of abbreviations

CK, Chapman-Kolmogorov; HMM, hidden Markov model; indel, insertion/deletion; LHS, local history set; MSA, multiple sequence alignment; PAS, preserved ancestral site; PWA, pairwise alignment.

Table of contents

Introduction **pp.3-5**
Results & Discussion **pp.6-31**
1. Perturbation approximations of probabilities **pp.6-22**
1.1. Perturbation expansion of multiplication factor contributed from local indel histories pp.6-8
1.2. Perturbation analyses on local PWA pp.8-17
1.3. Perturbation analyses on local MSA pp.17-22
2. Goodness of approximation by past indel models **pp.23-31**
2.1. Geometric indel length distributions pp.23-24
2.2. Models pursuing biological realism pp.24-29
2.3. Violation of Chapman-Kolmogorov equation pp.29-31
Conclusions **p.32**
Methods **p.32**
Authors' contributions **p.32**
Acknowledgements **p. 33**
Appendix **pp.34-55**
A1. Perturbation calculation of multiplication factors for PWAs between ancestral and decendant sequences pp.34-47
A.1.1. For case (iii) local PWAs pp.34-35
A.1.2. For case (iv) local PWAs pp.35-45
A.1.3. System of integral equations providing "exact solutions" for case (i) & (iii) local PWAs pp.45-47
A2. Perturbation calculation of multiplication factors for MSAs pp.47-55
Tables 1-4 **pp.56-59**
Figures 1-7 (with legends) **pp.60-67**
References **pp.68-69**

Introduction

The evolution of DNA, RNA, and protein sequences is driven by mutations such as base substitutions, insertions and deletions (indels), recombination, and other genomic rearrangements (*e.g.*, Graur and Li 2000; Gascuel 2005; Lynch 2007). Thus far, analyses on substitutions have predominated in the field of molecular evolutionary study, in particular using the probabilistic (or likelihood) theory of substitutions that is now widely accepted (e.g., Felsenstein 1981, 2004; Yang 2006). However, some recent comparative genomic analyses have revealed that indels account for more base differences between the genomes of closely related species than substitutions (*e.g.*, Britten 2002; Britten et al. 2003; Kent *et al*. 2003; The International Chimpanzee Chromosome 22 Consortium 2004; The Chimpanzee Sequencing and Analysis Consortium 2005). It is therefore imperative to develop a stochastic model that enables us to reliably calculate the probability of sequence evolution via mutations including insertions and deletions.

Since the groundbreaking works by Bishop and Thompson (1986) and by Thorne, Kishino and Felsenstein (1991), many studies have been done to develop and apply methods to calculate the probabilities of pairwise alignments (PWAs) and multiple sequence alignments (MSAs) under the probabilistic models aiming to incorporate the effects of indels, and such methods have greatly improved in terms of the computational efficiency and the scope of application. See excellent reviews for details on this topic (e.g., Rivas 2005; Bradley and Holmes 2007; Miklós et al. 2009). A majority of these studies are based on hidden Markov models (HMMs) or transducer theories. Both of them calculate the indel component of an alignment probability as a product of inter-column transition probabilities or of block-wise contributions. Unfortunately, they have two fundamental problems, one regarding the theoretical grounds and the other regarding the biological realism. Regarding the theoretical grounds, it is unclear whether or not a HMM or a transducer is related with any *genuine* evolutionary model, which describes the evolution of an *entire* sequence via indels along the time axis. Regarding the biological realism, the standard HMMs or transducers can at best handle geometric distributions of indel lengths, whereas many empirical studies showed that the real indel lengths are distributed according to the power-law *e.g.*, Gonnet et al. 1992; Benner et al. 1993; Gu and Li 1995; Kent et al. 2003; Zhang and Gerstein 2003; Chang and Benner 2004; The international Chimpanzee Chromosome 22 Consortium 2004; Yamane et al. 2006; Fan et al. 2007). Besides, it is very hard for the previous methods to incorporate the indel rate variation across regions, which were often observed empirically (*e.g.*, Gu et al. 2008). See the "background" section in part I (Ezawa, Graur and Landan 2015a) for more details on these problems.

In part I of this series of study (Ezawa, Graur and Landan 2015a), we established an *ab initio* formulation of a *genuine* stochastic evolutionary model, that is, a general continuous-time Markov model of sequence evolution via indels. Our evolutionary model allows any indel rate parameters including length distributions, but it does not impose any unnatural restrictions on indels. Thus, the model is naturally devoid of the aforementioned two problems. Aided by some techniques of the perturbation theory in physics (Dirac 1958; Messiah 1961a,b), we formally expanded the probability of an alignment into a series of terms with different numbers of indels. This expansion theoretically underpinned the stochastic evolutionary simulation method of Gillespie (1977). And we also showed that, if the indel model parameters satisfy a certain set of conditions, the *ab initio* probability of an alignment

is indeed factorable into the product of an overall factor and contributions from local alignments delimited by preserved ancestral sites (PASs). This result reconfirmed and generalized the conjecture that Miklós et al. (2004) made under their spatially and temporally homogeneous indel model.

In this paper, we turn our attention to more concrete problems. In other words, we focus on how to concretely calculate the contribution from each local alignment, assuming that the indel model satisfies the conditions for the factorability of alignment probabilities.

In Section 1 of Results & Discussion, we demonstrate how our *ab initio* perturbative formulation can be concretely used to approximately calculate the contributions to the alignment probabilities from regions separated by gapless columns (*i.e.*, local alignments). We examine all types of local pairwise alignments (PWAs) and some typical types of local multiple sequence alignments (MSAs). For each local alignment type, we calculate the fewest-indel contribution and the next-fewest-indel contribution to its probability. In this section, we also derive a system of integral equations that can be numerically solved to give "exact solutions" for some common types of local PWAs. Then, by comparing the fewest-indel contribution with the next-fewest-indel contribution, or with the "exact" solution, we examine the parameter region in which the fewest-indel contribution can approximate the alignment probability quite accurately. In Section 2 of Results & Discussion, we examine some representative models that were used in the past studies, including the HMM of Kim and Sinha (2007), in the light of our *ab initio* perturbative formulation. We attempt to delimit parameter regions where these models can safely approximate the *ab initio* probabilities under the *genuine* evolutionary model. The analyses will also suggest some modifications to these models that would enhance their reliability. And Appendix describes details on some perturbation calculations.

This paper is part II of a series of our papers that documents our efforts to develop, apply, and extend the *ab initio* perturbative formulation of the general continuous-time Markov model of sequence evolution via indels. Part I (Ezawa, Graur and Landan 2015a) gives the theoretical basis of this entire study. Part II (this paper) describes concrete perturbation calculations and examines the applicable ranges of other probabilistic models of indels. Part III (Ezawa, Graur and Landan 2015b) describes our algorithm to calculate the first approximation of the probability of a given MSA and simulation analyses to validate the algorithm. Finally, part IV (Ezawa, Graur and Landan 2015c) discusses how our formulation can incorporate substitutions and other mutations, such as duplications and inversions.

This paper basically uses the same conventions as used in part I (Ezawa, Graur and Landan 2015a). Briefly, a sequence state $s\,(\in S)$ is represented as an array of sites, each of which is either blank or equipped with some specific attributes. And each indel event is represented as an operator acting on the bra-vector, $\langle s|$, representing a sequence state. More specifically, the operator $\hat{M}_I(x,l)$ denotes the insertion of l sites between the x th and $(x+1)$ th sites, and the operator $\hat{M}_D(x_B,x_E)$ denotes the deletion of a sub-array between (and including) the x_B th and the x_E th sites. See Section 2 of part I for more details.

And, also as in part I, the following terminology is used. The term "an indel process" means a series of successive indel events with both the order and the specific timings specified, and the term "an indel history" means a series of successive indel events with only the order specified. And, throughout this paper, the union symbol,

such as in $A \bigcup B$ and $\bigcup_{i=1}^{I} A_i$, should be regarded as the union of *mutually disjoint* sets (*i.e.,* those satisfying $A \bigcap B = \varnothing$ and $A_i \bigcap A_j = \varnothing$ for $i \neq j \, (\in \{1,...,I\})$, respectively, where $\varnothing$ is an empty set), unless otherwise stated.

Results & Discussion

1. Perturbation approximations of probabilities

1.1. Perturbation expansion of multiplication factor contributed from local indel histories

In Section 4 of part I (Ezawa, Graur and Landan 2015a), we showed that the alignment probability can be factorized into the product of an overall factor and contributions from regions separated by gapless columns (as in Eqs.(4.1.8a,b) of part I for a PWA and Eqs.(4.2.9a,b,c) of part I for a MSA), if the following conditions are satisfied.

Condition (i): The indel rate parameters are independent of the portion of the sequence state outside of the region in question.

Condition (ii): The increment of the exit rate due to an indel event is independent of the portion of the sequence state outside of the region in question.

Condition (iii) (necessary only for MSAs): The probability of the root sequence state is factorable into the product of an overall factor and regional contributions. (This condition is represented as Eq.(4.2.8) of part I.)

Each regional contribution of the probability was symbolically represented as $\tilde{\mu}_P\left[\left(\tilde{\Lambda}^{ID}\left[\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$ (in Eq.(4.1.8b) of part I) for a PWA and $\tilde{\mathrm{M}}_P\left[\tilde{\Lambda}^{ID}_{\Psi}\left[C_{\mathrm{K}};\alpha[s_1,s_2,\ldots,s_{N^X}];T\right]\middle|T\right]$ (in Eq.(4.2.9c) of part I) for a MSA. (See below for the definitions of the symbols for the arguments.) Thus, as long as we can concretely calculate these regional contributions, we can obtain the specific value of the probability of a given alignment. However, because each of these factors is in general a summation of contributions from an infinite number of local indel histories, we usually need to approximate it by a summation over a finite number of histories. In this section, we will do this, again based on the perturbation expansion as unfolded in Section 3 of part I.

For a PWA, $\alpha(s^A,s^D)$, we decompose $\tilde{\Lambda}^{ID}\left[\gamma_\kappa;\alpha(s^A,s^D)\right]$, the set of local indel histories consistent with the local PWA confined in the region γ_κ, as: $\tilde{\Lambda}^{ID}\left[\gamma_\kappa;\alpha(s^A,s^D)\right]=\bigcup_{N_\kappa=N_{\min}\left[\gamma_\kappa;\alpha(s^A,s^D)\right]}^{\infty}\Lambda^{ID}\left[N_\kappa;\gamma_\kappa;\alpha(s^A,s^D)\right]$. Here $\Lambda^{ID}\left[N_\kappa;\gamma_\kappa;\alpha(s^A,s^D)\right]$ is the subset of PWA-consistent histories in γ_κ, each of which consists of N_κ indels, and $N_{\min}\left[\gamma_\kappa;\alpha(s^A,s^D)\right]$ is the minimum number of indels required to give rise to the local PWA of $\alpha(s^A,s^D)$ within γ_κ. Then, the "perturbation expansion" of the multiplication factor, $\tilde{\mu}_P\left[\left(\tilde{\Lambda}^{ID}\left[\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$, for the local PWA probability produced during time interval $[t_I,t_F]$, is given by:

$$\tilde{\mu}_P\left[\left(\tilde{\Lambda}^{ID}\left[\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\sum_{N_\kappa=N_{\min}\left[\gamma_\kappa;\alpha(s^A,s^D)\right]}^{\infty}\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa;\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]. \quad \text{--- Eq.(1.1.1a)}$$

Here

$$\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa;\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$\equiv \sum_{\bar{\hat{M}}[\gamma_\kappa]\in\Lambda^{ID}\left[N_\kappa;\gamma_\kappa;\alpha(s^A,s^D)\right]} \mu_P\left[\left(\bar{\hat{M}}[\gamma_\kappa],[t_I,t_F]\right)\middle|(s^A,t_I)\right] \qquad \text{--- Eq.(1.1.1b)}$$

is the multiplication factor contributed from all PWA-consistent N_κ-event local indel histories.

For a MSA, $\alpha[s_1,s_2,...,s_{N^X}]$, we decompose $\tilde{\Lambda}^{ID}_\Psi\left[C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]$, the set of local indel histories along the tree (T) consistent with the local MSA confined in the region C_K, as:

$$\tilde{\Lambda}^{ID}_\Psi\left[C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]=\bigcup_{N_\mathrm{K}=N_\mathrm{min}\left[C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]}^{\infty}\Lambda^{ID}_\Psi\left[N_\mathrm{K};C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right].$$

Here, $\Lambda^{ID}_\Psi\left[N_\mathrm{K};C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]$ is the subset consisting of all MSA-consistent N_K-event local indel histories along the tree, and $N_\mathrm{min}\left[C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]$ is the minimum possible number of indels that can produce the sub-MSA within C_K. Using this, the "perturbation expansion" of the multiplication factor, $\tilde{\breve{\mathrm{M}}}_P\left[\tilde{\Lambda}^{ID}_\Psi\left[C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]\middle|T\right]$, is given by:

$$\tilde{\breve{\mathrm{M}}}_P\left[\tilde{\Lambda}^{ID}_\Psi\left[C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]\middle|T\right]$$
$$=\sum_{N_\mathrm{K}=N_\mathrm{min}\left[C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]}^{\infty}\tilde{\breve{\mathrm{M}}}_P\left[\Lambda^{ID}_\Psi\left[N_\mathrm{K};C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]\middle|T\right]. \qquad \text{---Eq.(1.1.2a)}$$

Here

$$\tilde{\breve{\mathrm{M}}}_P\left[\Lambda^{ID}_\Psi\left[N_\mathrm{K};C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]\middle|T\right]$$
$$\equiv \sum_{\left(s^{Root},\left\{\bar{\hat{M}}(b)\right\}_T[C_\mathrm{K}]\right)\in\Lambda^{ID}_\Psi\left[N_\mathrm{K};C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]}\left\{\begin{array}{l}\mu_P\left[s^{Root},s_0^{Root},n^{Root};C_\mathrm{K}\right]\\ \times\breve{\mathrm{M}}_P\left[\left\{\left[\bar{\hat{M}}(b)\right]_{LHS}\right\}_T[C_\mathrm{K}]\middle|\left(s^{Root},n^{Root}\right)\right]\end{array}\right\} \qquad \text{---Eq.(1.1.2b)}$$

is the multiplication factor contributed from all local histories in $\Lambda^{ID}_\Psi\left[N_\mathrm{K};C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]$.

Thus, the problems were reduced to those of how to enumerate the local indel histories in the subset, $\Lambda^{ID}\left[N_\kappa;\gamma_\kappa;\alpha(s^A,s^D)\right]$ (for a PWA) or $\Lambda^{ID}_\Psi\left[N_\mathrm{K};C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]$ (for a MSA), up to a particular number of events, N_κ or N_K, that was chosen to achieve a desired level of accuracy. Once the relevant local histories are enumerated, then, we can calculate the contribution from each of them according to the general formula. (The general formulas are in Eq.(4.1.1b) of part I accompanied by Eq.(3.1.8b) of part I for a local PWA, and in Eq.(4.2.6b) of part I accompanied by Eq.(4.2.4b) of part I for a local MSA.) In Subsections 1.2 and 1.3, we explicitly calculate the portions of the multiplication factors, $\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa;\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$ and $\tilde{\breve{\mathrm{M}}}_P\left[\Lambda^{ID}_\Psi\left[N_\mathrm{K};C_\mathrm{K};\alpha[s_1,s_2,...,s_{N^X}];T\right]\middle|T\right]$, for some typical gap-configurations of the

local PWAs and MSAs, respectively, that are contributed from the parsimonious local indel histories. Then we compare them with the portions contributed from the next-parsimonious local indel histories. Moreover, for a few simplest but commonest types of local PWAs, we also derive integral equation systems that give the "exact solutions" for the total multiplication factors, Eq.(1.1.1a). Then, we see how well the contributions from parsimonious indel histories alone can approximate the actual occurrence frequencies of the gap-configurations of local MSAs.

In this section, we assume that we are working with a model whose alignment probabilities are factorable, and that we are focusing on calculating the multiplication factor that comes from a single local alignment flanked by a pair of gapless columns (*i.e.*, a pair of PASs). We will mainly work in the state space $S = S^{II}$ (see Section 2 of part I for the definition of the space). This means that we will focus on calculating the probability of the *homology structure* of each local alignment (see, *e.g.*, Lunter et al. 2005). Let $\Delta L(s)$ be the number of sites that a sequence $s \in S^{II}$ has between the pair of PASs. We will re-assign the site numbers so that the left- and right-flanking PASs are numbered 1 and $\Delta L(s)+2$, respectively, and the sites in between them are numbered $2, ..., \Delta L(s)+1$. This will make it easy to apply the theory formulated thus far to the current situation. We will re-assign the ancestries $\upsilon(1) = L$ and $\upsilon(\Delta L(s)+2) = R$ to the left- and right-flanking PASs, respectively. And we will usually re-assign the ancestries $\upsilon(x) = x-1$ to the sites, $x = 2,..., \Delta L(s)+1$, of the ancestral sequence $s = s^A$ (or the root sequence $s = s^{Root}$ for a MSA) in between the PASs (see Figure 1 as an illustration).

In the following subsections, we will omit the symbol "$[\ \ ]_{LHS}$" that denoted a local history set (LHS) in part I. Because we will consider a PWA or MSA region that accommodates only one local history, the resulting LHS equivalence classes are trivial classes, each of which consists only of a single history. We will also omit the appended "$[C_{\mathrm{K}}]$" to indicate the MSA region when it is obvious.

1.2. Perturbation analyses on local PWA

In a PWA, $\alpha(s^A, s^D)$, a gap-configuration flanked by a pair of PASs corresponds to the portion of $\alpha(s^A, s^D)$ confined in γ_κ. If we are interested only in the homology relationships among sites (*i.e.*, homology structures (Lunter et al. 2005)), there are four broad types of gap-configurations (see Figure 2; see also Subsection 3.3 of part I (Ezawa, Graur and Landan 2015a) for complexities concerning this issue). (i) Neither the ancestral nor the descendant sequence has even a single site in between the pair of PASs (panel A of Figure 2). (ii) The ancestor has one or more site(s) but the descendant has no site in between the PASs (panel B). (iii) The ancestor has no site but the descendant has one or more sites in between (panel C). And (iv) both the ancestor and the descendant have one or more site(s) in between, but no ancestral site is related to any of the descendant sites (panel D). We will consider these cases in turn.

In case (i), the sequence states could be represented as $s^A = s^D = [L, R]$. In this case, $N_{\min}\left[\gamma_\kappa; \alpha(s^A, s^D)\right] = 0$, and thus there is only one fewest-indel history, [], where no indel event takes place. Therefore, in this case, the portion of the multiplication factor contributed by the fewest-indel history is:

$$\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = 0; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right] = 1. \quad \text{--- Eq.(1.2.1)}$$

In this case, there is no history consisting only of one indel event. And each next-fewest-indel history should be a two-event history of the form, $\left[\hat{M}_I(1,l), \hat{M}_D(2,l+1)\right]$ with $l = 1,\ldots,L^{CO}$, where $L^{CO} \equiv \min\{L_I^{CO}, L_D^{CO}\}$. Thus, the portion contributed by the next-fewest-indel histories is:

$$\begin{aligned}&\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = 2; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right] \\ &= \sum_{l=1}^{L^{CO}} \mu_P\left[\left(\left[\hat{M}_I(1,l), \hat{M}_D(2,l+1)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right].\end{aligned} \quad \text{--- Eq.(1.2.2a)}$$

Let $\langle s[l]| \equiv \langle s^A | \hat{M}_I(1,l)$ be the state resulting from the action of an insertion of l sites on s^A. Then, using Eq.(4.1.1b) and Eq.(3.1.8b) of part I, each summand is calculated as:

$$\begin{aligned}&\mu_P\left[\left(\left[\hat{M}_I(1,l), \hat{M}_D(2,l+1)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right] \\ &= \int_{t_I}^{t_F} dt \int_t^{t_F} dt' \left[\begin{array}{l} r_I(1,l; s^A, t)\, r_D(2,l+1; s[l], t') \\ \times \exp\left\{-\int_{t_I}^{t} d\tau\, \delta R_X^{ID}(s^A, s^A, \tau) - \int_t^{t'} d\tau\, \delta R_X^{ID}(s[l], s^A, \tau) - \int_{t'}^{t_F} d\tau\, \delta R_X^{ID}(s^D, s^A, \tau)\right\}\end{array}\right] \\ &= \int_{t_I}^{t_F} dt \int_t^{t_F} dt'\, r_I(1,l; s^A, t)\, r_D(2,l+1; s[l], t') \exp\left\{-\int_t^{t'} d\tau\, \delta R_X^{ID}(s[l], s^A, \tau)\right\}.\end{aligned}$$

--- Eq.(1.2.2b)

Here, as in Subsection 4.1 of part I, $\delta R_X^{ID}(s, s', t) \equiv R_X^{ID}(s,t) - R_X^{ID}(s',t)$ is the increment of the exit rate, which in turn is given by Eq.(2.4.1b') of part I. The second equation of Eq.(1.2.2b) follows from $\delta R_X^{ID}(s^A, s^A, \tau) = \delta R_X^{ID}(s^D, s^A, \tau) = 0$. We could at least numerically calculate Eq.(1.2.2b) once the specific functional forms of the indel rates and the exit rates are given. For example, in a space-time-homogeneous model, like Dawg's indel model (Cartwright 2005), Eqs.(2.4.4a,b) of part I, we have $\delta R_X^{ID}(s[l], s^A, \tau) = (\lambda_I + \lambda_D) l$, and Eq.(1.2.2b) is calculated as:

$$\begin{aligned}&\mu_P\left[\left(\left[\hat{M}_I(1,l), \hat{M}_D(2,l+1)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right] \\ &= \lambda_I f_I(l)\, \lambda_D f_D(l)\, \frac{\exp\left(-(\lambda_I + \lambda_D) l (t_F - t_I)\right) - 1 + (\lambda_I + \lambda_D) l (t_F - t_I)}{\left((\lambda_I + \lambda_D) l\right)^2}.\end{aligned} \quad \text{--- Eq.(1.2.2b')}$$

(Exactly the same expression can be obtained also from the "long indel" model, Eqs.(2.4.5a,b) of part I, if we notice the correspondence, Eqs.(2.4.7a,b,c,d) of part I.) Eq.(1.2.2b') (or Eq.(1.2.2b) itself) indicates that $\mu_P\left[\left(\left[\hat{M}_I(1,l), \hat{M}_D(2,l+1)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right] < \frac{1}{2} \lambda_I f_I(l)\, \lambda_D f_D(l)\, (t_F - t_I)^2$ for each $l = 1,\ldots,L^{CO}$. Applying this inequality to Eq.(1.2.2a) and using another inequality, $\sum_{l=1}^{L^{CO}} f_I(l)\, f_D(l) \le \sum_{l=1}^{L_I^{CO}} f_I(l) \sum_{l=1}^{L_D^{CO}} f_D(l) = 1$, we have:

$$\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = 2; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right] < \tfrac{1}{2} \lambda_I\, \lambda_D\, (t_F - t_I)^2. \quad \text{--- Eq.(1.2.3)}$$

Empirically, the rate of indels ($\lambda_I + \lambda_D$) is estimated to be at most on the order of 1/10 of the substitution rate (Lunter 2007; Cartwright 2009). Eq.(1.2.3) indicates that, in case (i), even if the elapsed time ($t_F - t_I$) is such that the substitution process is nearly

saturated, e.g., $\lambda_S(t_F - t_I) \approx 4$, where λ_S is the total substitution rate per site, the total contribution from the next-fewest-indel histories, Eq.(1.2.2a), is at most on the order of 1/10 of the contribution by the fewest-indel history, Eq.(1.2.1). Thus, we expect that the fewest-indel contribution should approximate the entire multiplication factor (Eq.(4.1.8b) in part I) very well in case (i).

Incidentally, taking advantage of the result for case (i), we can calculate the multiplication factor for a gapless PWA segment consisting of L^P (> 2) PASs. For clarity, let us consider an indel model as discussed in Subsection V-1. Then, in each of the $L^P - 1$ inter-PAS positions this segment contains, a null indel history could occur independently of those in other positions. Let $\upsilon_1^P, \upsilon_2^P, ..., \upsilon_{L^P}^P$ be the ancestries assigned to the L^P PASs in this segment (in spatial order). Focusing on the segment, we have $s^A = s^D = \left[\upsilon_1^P, \upsilon_2^P, ..., \upsilon_{L^P}^P\right]$. Then, excluding the contributions from both ends, the total multiplication factor for this segment is expressed as:

$$\prod_{i=1}^{L^P-1} \tilde{\mu}_P\left[\left(\tilde{\Lambda}^{ID}\left[\ \gamma_{\kappa_i}; \alpha(s^A = [\upsilon_i^P, \upsilon_{i+1}^P], s^D = [\upsilon_i^P, \upsilon_{i+1}^P])\right], [t_I, t_F]\right) \middle| \ (s^A = [\upsilon_i^P, \upsilon_{i+1}^P], t_I)\right].$$

Here $\tilde{\mu}_P\left[\left(\tilde{\Lambda}^{ID}\left[\ \gamma_{\kappa_i}; \alpha(s^A = [\upsilon_i^P, \upsilon_{i+1}^P], s^D = [\upsilon_i^P, \upsilon_{i+1}^P])\right], [t_I, t_F]\right) \middle| \ (s^A = [\upsilon_i^P, \upsilon_{i+1}^P], t_I)\right]$ is the total multiplication factor for case (i) with $L = \upsilon_i^P$ and $R = \upsilon_{i+1}^P$. Under a space-homogeneous model, we can further simplify the above product as:

$\left(\tilde{\mu}_P\left[\left(\tilde{\Lambda}^{ID}\left[\ \gamma_\kappa; \alpha(s^A = [L,R], s^D = [L,R])\right], [t_I, t_F]\right) \middle| \ (s^A = [L,R], t_I)\right]\right)^{L^P-1}$. Here we used the uniform multiplication factor,

$\tilde{\mu}_P\left[\left(\tilde{\Lambda}^{ID}\left[\ \gamma_\kappa; \alpha(s^A = [L,R], s^D = [L,R])\right], [t_I, t_F]\right) \middle| \ (s^A = [L,R], t_I)\right]$.

In case (ii), we assume that the ancestral state has ΔL^A sites in between the flanking PASs. Thus, the ancestral and descendant states could be represented as $s^A = \left[L, 1, ..., \Delta L^A, R\right]$ and $s^D = \left[L, R\right]$, respectively. As long as $\Delta L^A \le L_D^{CO}$, $N_{\min}\left[\gamma_\kappa; \alpha(s^A, s^D)\right] = 1$, and there is only one fewest-indel history, $\left[\hat{M}_D(2, \Delta L^A + 1)\right]$, which consists of a single event that deletes the ancestral sites in between the PASs. Therefore, the contribution by the fewest-indel history is:

$$\begin{aligned}
&\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = 1; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| \ (s^A, t_I)\right] \\
&= \int_{t_I}^{t_F} dt\ r_D(2, \Delta L^A + 1; s^A, t) \exp\left\{-\int_{t_I}^{t} d\tau\ \delta R_X^{ID}(s^A, s^A, t) - \int_{t}^{t_F} d\tau\ \delta R_X^{ID}(s^D, s^A, t)\right\} \\
&= \int_{t_I}^{t_F} dt\ r_D(2, \Delta L^A + 1; s^A, t) \exp\left\{-\int_{t}^{t_F} d\tau\ \delta R_X^{ID}(s^D, s^A, t)\right\}.
\end{aligned}$$

--- Eq.(1.2.4)

Each next-fewest indel history is composed of two indel events. There are two types. (a) Two successive deletions, $\left[\hat{M}_D(x, x + l - 1), \hat{M}_D(2, \Delta L^A - l + 1)\right]$ with $l = 1, ..., \Delta L^A - 1$ and $x = 2, ..., \Delta L^A - l + 2$. And (b) an insertion followed by a deletion, $\left[\hat{M}_I(x, l), \hat{M}_D(2, \Delta L^A + l + 1)\right]$ with $l = 1, ..., \min\{L_I^{CO}, L_D^{CO} - \Delta L^A\}$ and $x = 1, ..., \Delta L^A + 1$. Thus, in this case, the total contribution by the next-fewest indel histories is given by:

$\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = 2; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| \ (s^A, t_I)\right] = \tilde{\mu}_P[(a)] + \tilde{\mu}_P[(b)]$. --- Eq.(1.2.5a)

Here,

$$\tilde{\mu}_P[(a)] \equiv \sum_{l=1}^{\Delta L^A-1} \sum_{x=2}^{\Delta L^A-l+2} \mu_P\left[\left(\left[\hat{M}_D(x,x+l-1), \hat{M}_D(2,\Delta L^A-l+1)\right],[t_I,t_F]\right) \middle| (s^A,t_I)\right]$$

--- Eq.(1.2.5b)

is the sum of contributions from the histories of type (a). And

$$\tilde{\mu}_P[(b)] \equiv \sum_{l=1}^{\min\{L_I^{CO},L_D^{CO}-\Delta L^A\}} \sum_{x=1}^{\Delta L^A+1} \mu_P\left[\left(\left[\hat{M}_I(x,l), \hat{M}_D(2,\Delta L^A+l+1)\right],[t_I,t_F]\right) \middle| (s^A,t_I)\right]$$

--- Eq.(1.2.5c)

is the sum of contributions from the histories of type (b). Let $\left\langle s^A \cdot [x,-l]\right| \equiv \left\langle s^A \middle| \hat{M}_D(x,x+l-1)$ be the intermediate state in each type (a) history. Then, the history's contribution is calculated as:

$$\mu_P\left[\left(\left[\hat{M}_D(x,x+l-1), \hat{M}_D(2,\Delta L^A-l+1)\right],[t_I,t_F]\right) \middle| (s^A,t_I)\right]$$

$$= \int_{t_I}^{t_F} dt \int_t^{t_F} dt' \begin{bmatrix} r_D(x,x+l-1;s^A,t)\, r_D(2,\Delta L^A-l+1;s^A\cdot[x,-l],t') \\ \times \exp\left\{-\int_t^{t'} d\tau\, \delta R_X^{ID}(s^A\cdot[x,-l],s^A,\tau) - \int_{t'}^{t_F} d\tau\, \delta R_X^{ID}(s^D,s^A,\tau)\right\} \end{bmatrix}.$$

--- Eq.(1.2.5d)

Similarly, let $\left\langle s^A \cdot [x,+l]\right| \equiv \left\langle s^A \middle| \hat{M}_I(x,l)$ be the intermediate state in each type (b) history. Then, the history's contribution is calculated as:

$$\mu_P\left[\left(\left[\hat{M}_I(x,l), \hat{M}_D(2,\Delta L^A+l+1)\right],[t_I,t_F]\right) \middle| (s^A,t_I)\right]$$

$$= \int_{t_I}^{t_F} dt \int_t^{t_F} dt' \begin{bmatrix} r_I(x,l;s^A,t)\, r_D(2,\Delta L^A+l+1;s^A\cdot[x,+l],t') \\ \times \exp\left\{-\int_t^{t'} d\tau\, \delta R_X^{ID}(s^A\cdot[x,+l],s^A,\tau) - \int_{t'}^{t_F} d\tau\, \delta R_X^{ID}(s^D,s^A,\tau)\right\} \end{bmatrix}.$$

--- Eq.(1.2.5e)

Eq.(1.2.4) and Eqs.(1.2.5a-e) can indeed be calculated at least numerically once the specific functional forms of the indel rates and the exit rates are given. For example, under Dawg's indel model (Eqs.(2.4.4a,b) in part I) or the "long indel" model as noted in case (i), we have $\delta R_X^{ID}(s^D,s^A,\tau) = -(\lambda_I+\lambda_D)\Delta L^A$, and Eq.(1.2.4) becomes:

$$\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa=1;\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right) \middle| (s^A,t_I)\right]$$

$$= \lambda_D f_D(\Delta L^A)\, \frac{\exp\left((\lambda_I+\lambda_D)\Delta L^A(t_F-t_I)\right)-1}{(\lambda_I+\lambda_D)\Delta L^A}.$$

--- Eq.(1.2.4')

Similarly, using $\delta R_X^{ID}(s^A\cdot[x,-l],s^A,\tau) = -(\lambda_I+\lambda_D)l$ and $\delta R_X^{ID}(s^A\cdot[x,+l],s^A,\tau) = +(\lambda_I+\lambda_D)l$, Eqs.(1.2.5d,e) in Dawg's model (and also in the "long indel" model) are calculated, respectively, as:

$$\mu_P\left[\left(\left[\hat{M}_D(x,x+l-1), \hat{M}_D(2,\Delta L^A-l+1)\right],[t_I,t_F]\right) \middle| (s^A,t_I)\right]$$

$$= \frac{\lambda_D f_D(l)\, \lambda_D f_D(\Delta L^A-l)}{(\lambda_I+\lambda_D)(\Delta L^A-l)} \left[\frac{e^{(\lambda_I+\lambda_D)\Delta L^A(t_F-t_I)}-1}{(\lambda_I+\lambda_D)\Delta L^A} - \frac{e^{(\lambda_I+\lambda_D)l(t_F-t_I)}-1}{(\lambda_I+\lambda_D)l}\right],$$

--- Eq.(1.2.5d')

$$\mu_P\left[\left(\left[\hat{M}_I(x,l),\hat{M}_D(2,\Delta L^A+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$

$$=\frac{\lambda_I f_I(l)\,\lambda_D f_D(\Delta L^A+l)}{(\lambda_I+\lambda_D)(\Delta L^A+l)}\left[\frac{e^{(\lambda_I+\lambda_D)\Delta L^A(t_F-t_I)}-1}{(\lambda_I+\lambda_D)\Delta L^A}-\frac{1-e^{-(\lambda_I+\lambda_D)l(t_F-t_I)}}{(\lambda_I+\lambda_D)l}\right]. \quad \text{--- Eq.(1.2.5e')}$$

Substituting Eqs.(1.2.5d',e') into Eqs.(1.2.5a,b,c), we can concretely calculate the total contribution of the next-fewest-indel histories. Figure 3 A shows the ratio of the total next-fewest-indel contribution, Eq.(1.2.5a), to the fewest-indel contribution, Eq.(1.2.4), as a function of the number of deleted ancestral sites (ΔL^A) and the expected number of indels per site ($(\lambda_I+\lambda_D)(t_F-t_I)$). For the figure, we used the following parameters: $f_I(l)=f_D(l)\propto l^{-1.6}$, $L_I^{CO}=L_D^{CO}=500$, and $\lambda_I/\lambda_D=1$. The power-law behavior of the indel length distributions broadly conforms to the past empirical observations *e.g.*, Gonnet et al. 1992; Benner et al. 1993; Gu and Li 1995; Kent et al. 2003; Zhang and Gerstein 2003; Chang and Benner 2004; Yamane et al. 2006; Fan et al. 2007). Here, the exact overall indel rate ($\lambda_I+\lambda_D$) does not matter, because the probabilities are invariant under the simultaneous rescaling of the rate and the time interval (t_F-t_I) that keeps the expected number of indels per site (*i.e.*, $(\lambda_I+\lambda_D)(t_F-t_I)$) unchanged. Moreover, we confirmed that the results remain almost the same even if we use $L_I^{CO}=L_D^{CO}=5000$. Now, as indicated by Figure 3A, each curve for a fixed $(\lambda_I+\lambda_D)(t_F-t_I)$ reaches an asymptotic value slightly above 1 when ΔL^A is sufficiently large. Thus, to define a threshold within which the fewest-indel histories alone are likely to give a decent approximation of the probability, using the point at which the ratio is 1 (unity) is risky. Here, we tentatively define the threshold as the value of ΔL^A at which the ratio is 0.5. With this definition, the tentative threshold, $(\Delta L^A)_{0.5}^{(NP-ii)}$, is around 128, 31, 12, 6 and 3 when $(\lambda_I+\lambda_D)(t_F-t_I)$ is 0.01, 0.04, 0.1, 0.2 and 0.4, respectively (Table 1). Hence, we have a rough inversely proportional relationship: $(\Delta L^A)_{0.5}^{(NP-ii)}\approx 1.2/[(\lambda_I+\lambda_D)(t_F-t_I)]$, under the parameter setting used here.

In case (iii), we assume that the descendant state has ΔL^D sites in between the flanking PASs. Thus, the ancestral and descendant states could be represented as $s^A=[L,R]$ and $s^D=\left[L,\upsilon_1^D,\ldots,\upsilon_{\Delta L^D}^D,R\right]$, respectively. The ancestries satisfy $\upsilon_i^D\notin\{L,R\}$ for every $i=1,\ldots,\Delta L^D$, and $\upsilon_i^D\neq\upsilon_j^D$ for every pair (i,j) with $i\neq j$, and their details depend on the responsible indel history. (Actually, such details don't matter in the state space S^{II}, as explained in Subsection 2.1 of part I.) As long as $\Delta L^D\le L_I^{CO}$, $N_{\min}\left[\gamma_\kappa;\alpha(s^A,s^D)\right]=1$, and there is only one fewest-indel history, $\left[\hat{M}_I(1,\Delta L^D)\right]$. The history consists of a single event that inserts the descendant sites in between the PASs. As in case (ii), each next-fewest indel history is composed of two indel events, and classified into two types. (c) Two successive insertions, $\left[\hat{M}_I(1,\Delta L^D-l),\hat{M}_I(x,l)\right]$ with $l=1,\ldots,\Delta L^D-1$ and $x=1,\ldots,\Delta L^D-l+1$. And (d) an insertion followed by a deletion, $\left[\hat{M}_I(1,\Delta L^D+l),\hat{M}_D(x,x+l-1)\right]$ with $l=1,\ldots,\min\{L_D^{CO},L_I^{CO}-\Delta L^D\}$ and $x=2,\ldots,\Delta L^D+2$. The sum of contributions by the fewest-indel histories and that by the next-fewest-indel histories can be calculated as in case (ii). And their calculations are detailed in Appendix A1.1. If calculated under

the same setting as used for Figure 3 A and with the same value of $(\lambda_I + \lambda_D)(t_F - t_I)$, their ratio with $\Delta L^D = L$ is identical to that in case (ii) with $\Delta L^A = L$, because of the symmetry of the probabilities under the time reversal. Thus, the same conclusions can be drawn from Figure 3 A also on this case.

In case (iv), we assume that the ancestral and the descendant states have ΔL^A and ΔL^D sites, respectively, in between the flanking PASs. Thus, the ancestral and descendant states could be represented as $s^A = \left[L, 1, ..., \Delta L^A, R\right]$ and $s^D = \left[L, \upsilon_1^D, ..., \upsilon_{\Delta L^D}^D, R\right]$, respectively. Here, the descendant ancestries satisfy $\upsilon_i^D \notin \{L, 1, ..., \Delta L^A, R\}$ for every $i = 1, ..., \Delta L^D$, and $\upsilon_i^D \neq \upsilon_j^D$ for every pair (i, j) with $i \neq j$, and their details depend on the responsible indel history. (Again, the details don't matter in the space S^{II}.) As long as $\Delta L^A \leq L_D^{CO}$ and $\Delta L^D \leq L_I^{CO}$, $N_{\min}\left[\gamma_\kappa; \alpha(s^A, s^D)\right] = 2$. As indicated by Eqs.(A1.3c',d') in Appendix A1 of part I, there are three types of fewest-indel histories. (e) The deletion of the ancestral sites followed by an insertion of ΔL^D sites, $\left[\hat{M}_D(2, \Delta L^A + 1), \hat{M}_I(1, \Delta L^D)\right]$. (f) An insertion immediately on the right of the ancestral sites to be deleted, followed by the deletion, $\left[\hat{M}_I(\Delta L^A + 1, \Delta L^D + l), \hat{M}_D(2, \Delta L^A + l + 1)\right]$ with $l = 0, ..., \min\{L_I^{CO} - \Delta L^D, L_D^{CO} - \Delta L^A\}$. And (g) an insertion immediately on the left of the ancestral sites to be deleted, followed by the deletion, $\left[\hat{M}_I(1, \Delta L^D + l), \hat{M}_D(\Delta L^D + 2, \Delta L^A + \Delta L^D + l + 1)\right]$ also with $l = 0, ..., \min\{L_I^{CO} - \Delta L^D, L_D^{CO} - \Delta L^A\}$. In this case, each next-fewest-indel history is composed of three indel events, and classified into one of 6 broad types: (h) two successive deletions followed by an insertion; (i) a deletion, followed by an insertion, followed by a deletion; (j) an insertion followed by two successive deletions; (k) a deletion followed by two successive insertions; (l) an insertion, followed by a deletion, followed by an insertion; and (m) two successive insertions followed by a deletion. And these six broad types can be further sub-classified into 24 sub-types, as described in Appendix A1.2. The calculations of the sum of contributions from the fewest-indel local histories and that from the next-fewest-indel local histories are also detailed in Appendix A1.2. Table 2 shows their ratios calculated for some typical configurations under the same setting as for Figure 3 A. As the table indicates, the approximation by the fewest-indel histories alone is likely to be decent as long as the expected number of indels (*i.e.*, $(\lambda_I + \lambda_D)(t_F - t_I)$) and the gap lengths (i.e., ΔL^A and ΔL^D) are at most moderate.

It is difficult to calculate the summed contributions from local histories involving more indels, especially in case (iv). We could calculate the contribution from a single local history involving any number of indels if we use the algorithm for a "trajectory likelihood" given by Miklós et al. (2004). As we exemplified in Appendix A1.2, however, it is already quite hard to enumerate all possible local indel histories even up to the next-to-minimum number of involved indels. Nevertheless, if we consider only cases (i), (ii), and (iii) under a (locally) spatially homogeneous model, we can work out systems of "exact" integral equations that could in principle provide the numerical solutions for the total sum of contributions to a multiplication factor up to a desired level of accuracy (at some time and memory expenses). In the following, we derive a system of integral equations to give multiplication factors for

cases (i) and (ii). Another system of integral equations, which gives multiplication factors for cases (i) and (iii), is derived in Appendix A1.3.

Here, we assume that the indel rates are locally homogeneous, which means that the rates do not depend on the exact positions that the indels hit, *as long as* they are confined in the region that accommodates the local history. Thus, we assume that the indel rate is *locally* homogeneous and the exit rate is *locally* an affine function of the (local) sequence length, but that they may be non-homogeneous *globally*. (In terms of equations, we *locally* assume Eqs.(5.1.1a,b) of part I for the indel rates and Eq.(5.2.4) of part I for the *local* exit rate, but we assume something like Eqs.(5.3.2a,b,c) of part I for the *global* exit rate.) We are now considering only cases (i) and (ii), in which (local) ancestral and descendant states should be $s^A = \left[L, 1, ..., \Delta L^A, R\right]$ and $s^D = \left[L, R\right]$, respectively, with $\Delta L^A = 0,1,2,...$. Because of the local homogeneity, Eqs.(V-1.1a,b), the exit rate $R_X^{ID}(s,t)$ of a state $s\ (\in S)$ in this context depends only on the (local) sequence length, $L(s) = 2 + \Delta L(s)$. Thus, $\Delta L(s)$ adequately represents the local sequence state s, and we let $R_X^{ID}(\Delta L(s), t)$ denote its (local) exit rate. The starting point of the equation system is the fundamental integral equation (Eq.(3.1.4) of part I) for the stochastic indel evolution operator $\hat{P}^{ID}(t_I, t_F)$. We sandwich the fundamental integral equation with $\left\langle s^A \right|$ and $\left| s^D \right\rangle$, and expand the instantaneous mutation operator $\hat{Q}_M^{ID}(t)$ using its definition (*i.e.*, Eq.(3.1.1c) of part I supplemented by Eqs.(2.4.2b',c') of part I). Because we know that the flanking PASs, which are assigned the ancestries L and R, have never been struck by any indels, we can ignore the effects of indels that hit the PASs. And, because we are now focusing on the local alignment, we will also ignore the indels completely outside of the region delimited by the PASs. Thus, we have:

$$\begin{aligned}
\left\langle s^A \middle| \hat{P}^{ID}(t_I, t_F) \middle| s^D \right\rangle &= \left\langle s^A \middle| s^D \right\rangle \exp\left\{ -\int_{t_I}^{t_F} dt\, R_X^{ID}\left(\Delta L^A, t\right) \right\} \\
&+ \sum_{x=1}^{\Delta L^A+1} \sum_{l=1}^{L_I^{CO}} \int_{t_I}^{t_F} dt \left[\exp\left\{ -\int_{t_I}^{t} d\tau\, R_X^{ID}\left(\Delta L^A, \tau\right) \right\} g_I(l,t) \left\langle s^A \middle| \hat{M}_I(x,l) \hat{P}^{ID}(t, t_F) \middle| s^D \right\rangle \right] \\
&+ \sum_{l=1}^{\min\{\Delta L^A, L_D^{CO}\}} \sum_{x=2}^{\Delta L^A - l + 2} \int_{t_I}^{t_F} dt \left[\exp\left\{ -\int_{t_I}^{t} d\tau\, R_X^{ID}\left(\Delta L^A, \tau\right) \right\} g_D(l,t) \left\langle s^A \middle| \hat{M}_D(x, x+l-1) \hat{P}^{ID}(t, t_F) \middle| s^D \right\rangle \right].
\end{aligned}$$

--- Eq.(1.2.6)

In the present setting, the number of sites between the PASs, $\Delta L(s)$, uniquely determines the local sequence state s. Thus, we let the local states denoted as $\left\langle s^A \right| = \left\langle \Delta L^A \right|$, $\left\langle s^D \right| = \left\langle 0 \right|$, $\left\langle s^A \middle| \hat{M}_I(x,l) = \left\langle \Delta L^A + l \right|$, and $\left\langle s^A \middle| \hat{M}_D(x, x+l-1) = \left\langle \Delta L^A - l \right|$. We also introduce the notation, $P^{ID}\left(\Delta L \mapsto \Delta L'; [t, t']\right) \equiv \left\langle \Delta L \middle| \hat{P}^{ID}(t, t') \middle| \Delta L' \right\rangle$, for the probability that the local state with ΔL sites in between the PASs at time t became a state with $\Delta L'$ sites in between at time t'. Then, taking advantage of the independence of the indel rates, exit rates and $P^{ID}\left(\Delta L \mapsto \Delta L'; [t, t']\right)$ on the position of each indel (x), we have:

$$P^{ID}\left(\Delta L^A \mapsto 0;[t_I,t_F]\right) = \delta(\Delta L^A,0)\exp\left\{-\int_{t_I}^{t_F} dt\, R_X^{ID}\left(\Delta L^A=0,t\right)\right\}$$
$$+(\Delta L^A+1)\sum_{l=1}^{L_I^{CO}}\int_{t_I}^{t_F} dt\left[\exp\left\{-\int_{t_I}^{t} d\tau\, R_X^{ID}\left(\Delta L^A,\tau\right)\right\} g_I(l,t)\, P^{ID}\left(\Delta L^A+l \mapsto 0;[t,t_F]\right)\right]$$
$$+\sum_{l=1}^{\min\{\Delta L^A, L_D^{CO}\}}(\Delta L^A-l+1)\int_{t_I}^{t_F} dt\left[\exp\left\{-\int_{t_I}^{t} d\tau\, R_X^{ID}\left(\Delta L^A,\tau\right)\right\} g_D(l,t)\, P^{ID}\left(\Delta L^A-l \mapsto 0;[t,t_F]\right)\right].$$

--- Eq.(1.2.7)

Here $\delta(\Delta L,\Delta L')$ is Kronecker's delta, which equals 1 if $\Delta L=\Delta L'$, and 0 if $\Delta L\neq\Delta L'$. Eq.(1.2.7) gives the desired system of integral equations for the "exact" probabilities, $P^{ID}\left(\Delta L^A \mapsto 0;[t_I,t_F]\right)$, with non-negative integers $\Delta L^A = 0,1,2,\ldots$. This equation holds for every non-negative integer ΔL^A and even if we replace the initial time t_I with any time in the closed interval $[t_I,t_F]$. Thus, the equations can be solved iteratively, starting with the "zero-event approximation" of the probability, $P_{\langle 0\rangle}^{ID}\left(\Delta L^A \mapsto 0;[t,t_F]\right) = \delta(\Delta L^A,0)\exp\left\{-\int_{t}^{t_F} d\tau\, R_X^{ID}\left(\Delta L^A=0,\tau\right)\right\}$, and calculating the approximation at the n_S th step, $P_{\langle n_S\rangle}^{ID}\left(\Delta L^A \mapsto 0;[t,t_F]\right)$, from the approximation at the previous step via the recursion relation:

$$P_{\langle n_S\rangle}^{ID}\left(\Delta L^A \mapsto 0;[t,t_F]\right) = \delta(\Delta L^A,0)\exp\left\{-\int_{t}^{t_F} d\tau\, R_X^{ID}\left(\Delta L^A=0,\tau\right)\right\}$$
$$+(\Delta L^A+1)\sum_{l=1}^{L_I^{CO}}\int_{t}^{t_F} dt'\left[\exp\left\{-\int_{t}^{t'} d\tau\, R_X^{ID}\left(\Delta L^A,\tau\right)\right\} g_I(l,t')\, P_{\langle n_S-1\rangle}^{ID}\left(\Delta L^A+l \mapsto 0;[t',t_F]\right)\right]$$
$$+\sum_{l=1}^{\min\{\Delta L^A, L_D^{CO}\}}(\Delta L^A-l+1)\int_{t}^{t_F} dt'\left[\exp\left\{-\int_{t}^{t'} d\tau\, R_X^{ID}\left(\Delta L^A,\tau\right)\right\} g_D(l,t')\, P_{\langle n_S-1\rangle}^{ID}\left(\Delta L^A-l \mapsto 0;[t',t_F]\right)\right].$$

--- Eq.(1.2.7')

If N_{ID} iteration steps are performed, the resulting probability, $P_{\langle N_{ID}\rangle}^{ID}\left(\Delta L^A \mapsto 0;[t_I,t_F]\right)$, is the summation of the probabilities over all possibly responsible local histories consisting of up to (and including) N_{ID} indel events. After the iteration is finished, the multiplication factor will be obtained by following its definition (*i.e.*, Eq.(4.1.1b) in part I). We have:

$$\begin{aligned}
&\tilde{\mu}_P^{\langle N_{ID}\rangle}\left[\left(\tilde{\Lambda}^{ID}\left[\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right] \\
&\equiv \sum_{N_\kappa=0}^{N_{ID}}\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa;\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right] \\
&= P_{\langle N_{ID}\rangle}^{ID}\left(\Delta L^A \mapsto 0;[t_I,t_F]\right)\Big/ P\left[\left([],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right] \\
&= \exp\left\{+\int_{t_I}^{t_F} dt\, R_X^{ID}\left(\Delta L^A,t\right)\right\} P_{\langle N_{ID}\rangle}^{ID}\left(\Delta L^A \mapsto 0;[t_I,t_F]\right).
\end{aligned}$$

--- Eq.(1.2.8)

The accuracy of the numerical solutions will depend on how finely we partition the time interval $[t_I,t_F]$. If the interval is partitioned into N_P equal-sized sub-intervals, we could in principle achieve an accuracy of $O\left((N_P)^{-4}\right)$ under Simpson's rule (*e.g.*,

Press et al. 1992). However, as the number of sub-intervals increases, it would take longer to complete the calculation. A naïve implementation of the aforementioned numerical iteration would have the time complexity of $O\left(N_{ID}(L^{CO})^2(N_P)^2\right)$ and the space complexity of $O\left(L^{CO}N_P\right)$, when we want to obtain the total probabilities of local histories composed of up to N_{ID} indels each, with $\Delta L^A = 0,1,2,\ldots,\Delta L^A_{\max}(\leq L^{CO})$. Here we set $L_I^{CO} = L_D^{CO} \equiv L^{CO}$. This becomes impractically slow when either L^{CO} or N_P is large, *e.g.*, around 1000 or greater. It is likely that N_P does not have to be this large, as it would be usually enough to set N_P around 100 or smaller. However, L^{CO} will often be around 1000 or greater, indeed making the naïve algorithm too slow to be practical. Fortunately, we can avoid this problem by rewriting the recursion equation, Eq.(1.2.7'), as:

$$P^{ID}_{\langle n_S\rangle}\left(\Delta L^A \mapsto 0;[t,t_F]\right) = \delta(\Delta L^A,0)\exp\left\{-\int_t^{t_F} d\tau\, R_X^{ID}\left(\Delta L^A = 0,\tau\right)\right\} + \int_t^{t_F} dt'\left[\exp\left\{-\int_t^{t'} d\tau\, R_X^{ID}\left(\Delta L^A,\tau\right)\right\}\Phi^{ID}_{\langle n_S\rangle}\left(\Delta L^A \mapsto 0;[t',t_F]\right)\right]. \quad \text{--- Eq.(1.2.9a)}$$

Here, the "auxiliary function," $\Phi^{ID}_{\langle n_S\rangle}\left(\Delta L^A \mapsto 0;[t',t_F]\right)$ is given by:

$$\Phi^{ID}_{\langle n_S\rangle}\left(\Delta L^A \mapsto 0;[t,t_F]\right) \equiv (\Delta L^A+1)\sum_{l=1}^{L_I^{CO}}\left[g_I(l,t)\,P^{ID}_{\langle n_S-1\rangle}\left(\Delta L^A + l \mapsto 0;[t,t_F]\right)\right] + \sum_{l=1}^{\min\{\Delta L^A, L_D^{CO}\}}\left[(\Delta L^A - l + 1)g_D(l,t)\,P^{ID}_{\langle n_S-1\rangle}\left(\Delta L^A - l \mapsto 0;[t,t_F]\right)\right]. \quad \text{--- Eq.(1.2.9b)}$$

Consider the following "two-sub-step" algorithm. In the first sub-step (in each iteration step), it calculates $\Phi^{ID}_{\langle n_S\rangle}\left(\Delta L^A \mapsto 0;[t,t_F]\right)$'s via Eq.(1.2.9b) and stores them for all $\Delta L^A = 0,1,2,\ldots,L^{CO}$ at all time points, $t = t_I + i\,\frac{t_F - t_I}{N_P}$ with $i = 0,1,\ldots,N_P$. And, in the second sub-step, it uses them to calculate the probabilities $P^{ID}_{\langle n_S\rangle}\left(\Delta L^A \mapsto 0;[t,t_F]\right)$ via Eq.(1.2.9a) for the same values of ΔL^A and t. This algorithm can reduce the time-complexity to $O\left(N_{ID}\,L^{CO}(L^{CO}+N_P)N_P\right)$ while keeping the space complexity to be $O\left(L^{CO}N_P\right)$. This algorithm does finish in a practical amount of time (typically on the order of an hour or shorter when implemented in Perl). But it may still be too slow to perform each time we evaluate the probability of the gap configuration of a local MSA. Good news is that a single run of the iteration algorithm *inevitably* calculates the probabilities for all $\Delta L^A = 0,1,2,\ldots,\Delta L^A_{\max}(\leq L^{CO})$ at all temporal partitioning points, $t = t_I + i\,\frac{t_F - t_I}{N_P}$ ($i = 1,\ldots,N_P - 1$), as well as at $t = t_I$ and $t = t_F$. Thus, once we calculate the probabilities with a fixed set of model parameters, we could use them to calculate the probabilities of various alignments (under various phylogenetic trees), as long as the model parameters remain unchanged. In any case, the time and space complexities might be further reduced without considerably compromising the accuracy by a clever beforehand discarding of terms that are unlikely to make significant contributions to the final probabilities. Panels B and C of Figure 3 show the ratios of the multiplication factors, Eq.(1.2.8) at $N_{ID} = 1, 2, 5, 10, 20$ iteration steps, to that at $N_{ID} = 50$ steps.

They were calculated with $\Delta L^A_{\max} = 300$ and $(\lambda_I + \lambda_D)(t_F - t_I)/N_P = 0.002$, and under the same setting as that used for Figure 3 A. (We confirmed that results remain almost the same even if L^{CO} is greater than 1000, instead of $L^{CO} = 500$ used for the figure.) When $N_{ID} \geq 2$, we actually started from $N_{ID} = 2$, at which the probabilities were calculated using Eqs.(1.2.2a,b'), Eq.(1.2.4') and Eqs.(1.2.5a,b,c,d',e'), in stead of from $N_{ID} = 0$ as mentioned above, in order to enhance the accuracy of the approximation. As indicated by the panels B and C, the accuracy of the probabilities improves in a step-wise manner as the number of iterations increases. Now that the "exact" probabilities are available in cases (i) and (ii), we can use them to define more precise threshold values of ΔL^A within which the approximate probabilities are expected to be quite accurate. For example, let $(\Delta L^A)^{(1-ii)}_{0.5}$ be the value of ΔL^A at which the approximation by the 1-event local indel history alone accounts for 50% of the total probability of the local PWA. As shown in Table 1, $(\Delta L^A)^{(1-ii)}_{0.5}$ is slightly larger than the aforementioned "tentative threshold," $(\Delta L^A)^{(NP-ii)}_{0.5}$. Indeed, we can observe a rough inversely proportional relationship, $(\Delta L^A)^{(1-ii)}_{0.5} \approx 1.6/[(\lambda_I + \lambda_D)(t_F - t_I)]$. Thus, we expect that the tentative threshold ($(\Delta L^A)^{(NP-ii)}_{0.5}$) should work as a slightly conservative criterion for the decency of the approximation by the fewest-indel histories alone. We could also define the threshold, $(\Delta L^A)^{(N_{ID}-ii)}_{0.5}$, at which the probability calculated up to (and including) N_{ID} iterations, *i.e.*, $P^{ID}_{\langle N_{ID}\rangle}\left(\Delta L^A \mapsto 0; [t, t_F]\right)$, is 1/2 (=0.5) of the "exact solution." As indicated by Table 1, $(\Delta L^A)^{(N_{ID}-ii)}_{0.5}$ is over N_{ID} times as large as $(\Delta L^A)^{(1-ii)}_{0.5}$. This implies that, even when the fewest-indel histories alone give a poor approximation of the probability, taking account of the next-fewest-indel histories, or of the histories involving yet more indels, will considerably expand the range of ΔL^A where the approximate probability is quite accurate.

Following the similar procedures, this time starting from the integral equation, Eq.(III-1.2), we can also derive a system of integral equations for the multiplication factors for cases (i) and (iii), as described in Appendix A1.3. Thanks to the symmetry of the probabilities under the time reversal, panels B and C of Figure 3 can also be interpreted as the results of numerical calculations of this equation system, under the same setting as above except that $\Delta L^A_{\max} = 300$ is replaced by $\Delta L^D_{\max} = 300$.

1.3. Perturbation analyses on local MSA

Compared to contributions to local PWAs, those to local MSAs are much more complex. In this subsection, we consider some simple but common patterns, under a tree T with three OTUs, corresponding to the external nodes, n_1, n_2 and n_3. Here, we regard its single internal node as the root node n^{Root} for simplicity (panel A of Figure 4). Let b_m ($m = 1,2,3$) be the branch that connects the nodes n^{Root} and n_m. Let $s_m \in S$ ($m = 1,2,3$) be the (local) sequence state at node n_m. Then, we consider the gap-configurations of the MSAs of the three sequences, $\alpha[s_1, s_2, s_3]$, as well as the consistent sequence states $s^{Root} \in S$ at the root node n^{Root}. As in the previous subsections, we focus on the portion of MSAs delimited by a pair of PASs, whose ancestries are denoted as L and R. Here we consider four typical situations (see

Figure 4; see also Subsection 3.4 of part I (Ezawa, Graur and Landan 2015a) for complexities concerning this issue). (I) None of $\{s_1, s_2, s_3\}$ have any site in between the PASs (Figure 4, panel B). (II) s_1 and s_2 share the identical set of sites in between the PASs, but s_3 has no site in between (panel C). (III) s_1 has a set of sites in between the PASs, but neither s_2 nor s_3 has even a single site in between (panel D). And (IV) s_1 has a set of sites in between the PASs, but s_3 has no site in between, and s_2 lacks a run of some, but not all, of contiguous sites of s_1 in between the PASs (panel E). These situations are not restricted to the 3-OTU trees but widely applicable to each portion surrounding a trivalent node of any tree topology, although they never exhaust all gap configurations. The time at n^{Root} will be represented as t_I, and the time at node n_m will be represented as $t_{F:m}$. The indel parameters along branch b_m will be indicated by the subscript "$:m$."

Case (I) is represented by the external sequence states $s_1 = s_2 = s_3 = [L, R]$. In this case, we have $N_{\min}[C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T] = 0$. And the set of fewest-indel local histories, $\Lambda_{\Psi}^{ID}[N_{\mathrm{K}} = 0; C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T]$, is composed only of a no-indel history:

$$\left\{\vec{\hat{M}}(b_1) = \vec{\hat{M}}(b_2) = \vec{\hat{M}}(b_3) = [\ \ ]\right\}, \qquad \text{--- Eq.(1.3.1)}$$

with $s^{Root} = s_0^{Root} = [L, R]$. Thus, according to Eq.(1.1.2b), supplemented by Eqs.(4.2.4b,6b,8) of part I, the portion of the multiplication factor contributed by the fewest-indel local history is:

$$\tilde{\mathrm{M}}_P\left[\Lambda_{\Psi}^{ID}[N_{\mathrm{K}} = 0; C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T] \,\middle|\, T\right] = 1. \quad \text{--- Eq.(1.3.2)}$$

Here we used $\mu_P\left[s_0^{Root}, s_0^{Root}, n^{Root}; C_{\mathrm{K}}\right] = 1$. No single-event local history can result in the gap configuration in this case. Each next-fewest-indel history consists of two indels, and it can be represented as:

$$\left\{\vec{\hat{M}}(b_m) = \left[\hat{M}_I(1,l), \hat{M}_D(2,l+1)\right],\ \vec{\hat{M}}(b_{m'}) = [\ \ ]\ \ for\ {}^{\forall}m' \in \{1,2,3\} \setminus \{m\}\right\},$$

--- Eq.(1.3.3)

with $m \in \{1,2,3\}$, $l \in \left\{1,2,\ldots,L_{:m}^{CO}\ (\equiv \min\{L_{I:m}^{CO}, L_{D:m}^{CO}\})\right\}$, and with $s^{Root} = s_0^{Root}$ again. Thus, the total contribution to the multiplication factor by the next-fewest-indel histories can be calculated similarly to Eqs.(1.2.2a,b). We have:

$$\tilde{\mathrm{M}}_P\left[\Lambda_{\Psi}^{ID}[N_{\mathrm{K}} = 2; C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T] \,\middle|\, T\right]$$
$$= \sum_{m=1,2,3} \sum_{l=1}^{L^{CO}} \mu_P\left[\left(\left[\hat{M}_I(1,l), \hat{M}_D(2,l+1)\right], [t_I, t_{F:m}]\right) \,\middle|\, (s_0^{Root}, t_I)\right]\Bigg|_{b_m}. \qquad \text{--- Eq.(1.3.4a)}$$

Each summand can be calculated from Eq.(1.2.2b), by replacing s^A with s_0^{Root} and also replacing the time and rate parameters with those assigned to each branch. Especially, under Dawg's model, each summand is calculated as:

$$\mu_P\left[\left(\left[\hat{M}_I(1,l), \hat{M}_D(2,l+1)\right], [t_I, t_{F:m}]\right) \,\middle|\, (s_0^{Root}, t_I)\right]\Bigg|_{b_i}$$
$$= \lambda_{I:m} f_{I:m}(l)\, \lambda_{D:m} f_{D:m}(l)\, \frac{\exp\left(-(\lambda_{I:m} + \lambda_{D:m}) l (t_{F:m} - t_I)\right) - 1 + (\lambda_{I:m} + \lambda_{D:m}) l (t_{F:m} - t_I)}{\left((\lambda_{I:m} + \lambda_{D:m}) l\right)^2}.$$

--- Eq.(1.3.4b)

If the three branches share the same time interval and the indel rate parameters, the summed contribution from the next-fewest-indel histories, Eq.(1.3.4a), is exactly three times Eq.(1.2.2a) for a PWA. Indeed, this total contribution on a general tree can be calculated by summing Eq.(1.2.2a) (with appropriate parameters) over all branches of the tree. Following the same line of reasoning as around Eq.(1.2.3), this total contribution is roughly proportional to the summation of the squared branch lengths over all branches. This means that a richer sampling of the homologous sequences will not significantly increase, or might rather slightly decrease, this total contribution, as long as the maximum evolutionary distance remains at a similar level. Incidentally, any root sequence state of the type $s^{Root} = \left[L, 1, ..., \Delta L^{Root}, R\right]$ is also consistent with $\alpha[s_1, s_2, s_3]$ in this case. Such a state, however, would require at least three indels, in order to delete the extra sites, $1, ..., \Delta L^{Root}$, independently along the three branches. Thus, the contributions from the local indel histories with such root states would be smaller in general.

Case (II) is represented by the external sequence states $s_1 = s_2 = \left[L, 1, ..., \Delta L^{D12}, R\right]$ and $s_3 = \left[L, R\right]$. In this case, the "phylogenetic correctness" condition (see, *e.g.*, Chindelevitch et al. 2006; Section 3.4 of part I (Ezawa, Graur and Landan 2015a)) dictates that the root state s^{Root} must have all the sites with ancestries $1, ..., \Delta L^{D12}$. In this case, we have $N_{\min}\left[C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T\right] = 1$. And the set of fewest-indel local histories, $\Lambda_{\Psi}^{ID}\left[N_{\mathrm{K}} = 1; C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T\right]$, consists of a single element:

$$\left\{\vec{\hat{M}}(b_1) = \vec{\hat{M}}(b_2) = [\ \ ], \vec{\hat{M}}(b_3) = \left[\hat{M}_D(2, \Delta L^{D12} + 1)\right]\right\}, \quad \text{--- Eq.(1.3.5)}$$

with $s^{Root} = s_0^{Root} = \left[L, 1, ..., \Delta L^{D12}, R\right]$. Again, according to Eq.(1.1.2b), supplemented by Eqs.(4.2.4b,6b,8) of part I, the contribution by the fewest-indel local history turns out to be exactly the same as Eq.(1.2.4) for case (ii) of PWAs, with the parameters replaced with those assigned to the branch b_3, and with ΔL^A replaced with ΔL^{D12}. Especially, under Dawg's model, we have:

$$\begin{aligned} &\tilde{\mathrm{M}}_P\left[\Lambda_{\Psi}^{ID}\left[N_{\mathrm{K}} = 1; C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T\right] \middle| T\right] \\ &= \lambda_{D:3} f_{D:3}(\Delta L^{D12}) \frac{\exp\left((\lambda_{I:3} + \lambda_{D:3})\Delta L^{D12}(t_{F:3} - t_I)\right) - 1}{(\lambda_{I:3} + \lambda_{D:3})\Delta L^{D12}} . \end{aligned} \quad \text{--- Eq.(1.3.6)}$$

As in case (ii) of PWAs, each next-fewest-indel history is composed of two indel events, and there are two types of such histories. One is based on type (a) in case (ii):

$$\left\{\vec{\hat{M}}(b_1) = \vec{\hat{M}}(b_2) = [\ \ ], \vec{\hat{M}}(b_3) = \left[\hat{M}_D(x, x + l - 1), \hat{M}_D(2, \Delta L^{D12} - l + 1)\right]\right\}, \quad \text{--- Eq.(1.3.7a)}$$

with $l = 1, ..., \Delta L^{D12} - 1$, $x = 2, ..., \Delta L^{D12} - l + 2$, and also with $s^{Root} = s_0^{Root}$. And the other is based on type (b) in case (ii):

$$\left\{\vec{\hat{M}}(b_1) = \vec{\hat{M}}(b_2) = [\ \ ], \vec{\hat{M}}(b_3) = \left[\hat{M}_I(x, l), \hat{M}_D(2, \Delta L^{D12} + l + 1)\right]\right\}, \quad \text{--- Eq.(1.3.7b)}$$

with $l = 1, ..., \min\{L_{I:3}^{CO}, L_{D:3}^{CO} - \Delta L^{D12}\}$, $x = 1, ..., \Delta L^{D12} + 1$, and also with $s^{Root} = s_0^{Root}$ again. Thus the summed contributions over the next-fewest-indel histories is given by:

$$\tilde{\mathrm{M}}_P\left[\Lambda_{\Psi}^{ID}\left[N_{\mathrm{K}} = 2; C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T\right] \middle| T\right] = \tilde{\mathrm{M}}_P[(a)] + \tilde{\mathrm{M}}_P[(b)]. \quad \text{--- Eq.(1.3.8)}$$

Here $\tilde{\tilde{\mathrm{M}}}_P[(a)]$ and $\tilde{\tilde{\mathrm{M}}}_P[(b)]$ are given by exactly the same equations as Eqs.(1.2.5b,d) and Eqs.(1.2.5c,e), respectively, with the parameters replaced by those assigned to branch b_3, and with ΔL^A replaced by ΔL^{D12}. Especially, under Dawg's model, the contributions by the individual next-fewest-indel histories are given by Eqs.(1.2.5d',e') with an appropriate replacement of the parameters and the ancestral state. Thus, panel A of Figure 3 can also be interpreted exactly as the comparison between the fewest-indel history's contribution and the next-fewest-indel histories' total contribution in the current case. Incidentally, root sequences with some additional ancestral sites in between the PASs of $s_0^{Root} \equiv \left[L, 1, ..., \Delta L^{D12}, R\right]$ are also consistent with $\alpha[s_1, s_2, s_3]$ in this case. However, such root sequence states require at least three indels each to give rise to $\alpha[s_1, s_2, s_3]$. This is because the additional ancestral sites need to be deleted independently along b_1 and b_2, even if they are deleted simultaneously with the sites with the ancestries $1, ..., \Delta L^{D12}$ along b_3. Thus, in general, the indel histories consistent with such root states are expected to contribute much less to the multiplication factor.

Case (III) is represented by the external sequence states $s_1 = \left[L, 1, ..., \Delta L^{D1}, R\right]$ and $s_2 = s_3 = \left[L, R\right]$. In this case, the phylogenetic correctness condition does not require the root state to have any site in between the PASs. Thus we have $s_0^{Root} = \left[L, R\right]$. As in case (II), we have $N_{\min}\left[C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T\right] = 1$. And the set of fewest-indel local histories, $\Lambda_{\Psi}^{ID}\left[N_{\mathrm{K}} = 1; C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T\right]$, consists of a single element:

$$\left\{\bar{\hat{M}}(b_1) = \left[\hat{M}_I(1, \Delta L^{D1})\right], \bar{\hat{M}}(b_2) = \bar{\hat{M}}(b_3) = [\ \ ]\right\}, \quad \text{--- Eq.(1.3.9)}$$

with $s^{Root} = s_0^{Root}$. Again, as in case (II), the contribution by this local history turns out to be exactly the same as Eq.(A1.1.1) (in Appendix A1.1) for case (iii) of PWAs, with the parameters replaced with those assigned to the branch b_1, and with ΔL^D replaced with ΔL^{D1}. Especially, under Dawg's model, we have:

$$\tilde{\tilde{\mathrm{M}}}_P\left[\Lambda_{\Psi}^{ID}\left[N_{\mathrm{K}} = 1; C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T\right] \middle| T\right]$$
$$= \lambda_{I:1} f_{I:1}(\Delta L^{D1}) \frac{1 - \exp\left(-(\lambda_{I:1} + \lambda_{D:1})\Delta L^{D1}(t_{F:1} - t_I)\right)}{(\lambda_{I:1} + \lambda_{D:1})\Delta L^{D1}} . \quad \text{--- Eq.(1.3.10)}$$

As in case (iii) of PWAs, each next-fewest-indel history is composed of two indel events. Unlike case (iii) of PWAs, however, there are three types of such histories; two of them are similar to those in case (iii), but the other one is totally new. Specifically, the first one is based on type (c) in case (iii):

$$\left\{\bar{\hat{M}}(b_1) = \left[\hat{M}_I(1, \Delta L^{D1} - l), \hat{M}_I(x, l)\right], \bar{\hat{M}}(b_2) = \bar{\hat{M}}(b_3) = [\ \ ]\right\}, \quad \text{--- Eq.(1.3.11a)}$$

with $l = 1, ..., \Delta L^{D1} - 1$, $x = 1, ..., \Delta L^{D1} - l + 1$, and also with $s^{Root} = s_0^{Root}$. The second one is based on type (d) in case (iii):

$$\left\{\bar{\hat{M}}(b_1) = \left[\hat{M}_I(1, \Delta L^{D1} + l), \hat{M}_D(x, x + l - 1)\right], \bar{\hat{M}}(b_2) = \bar{\hat{M}}(b_3) = [\ \ ]\right\}, \quad \text{--- Eq.(1.3.11b)}$$

with $l = 1, ..., \min\{L_{D:1}^{CO}, L_{I:1}^{CO} - \Delta L^{D1}\}$, $x = 2, ..., \Delta L^{D1} + 2$, and also with $s^{Root} = s_0^{Root}$ again. The third one involves events along b_2 and b_3, instead of along b_1:

$$\left\{\bar{\hat{M}}(b_1)=[\ \ ],\bar{\hat{M}}(b_2)=\bar{\hat{M}}(b_3)=\left[\hat{M}_D(2,\Delta L^D+1)\right]\right\}. \qquad \text{--- Eq.(1.3.11c)}$$

It is consistent with the root state $s^{Root}=s_1=\left[L,1,...,\Delta L^{D1},R\right]$ instead of $s_0^{Root}=\left[L,R\right]$. It should be noted that there is only one local history of the third type. In this case, therefore, the summed contribution of the next-fewest-indel local histories is given by:

$$\tilde{\mathrm{M}}_P\left[\Lambda_\Psi^{ID}\left[N_\mathrm{K}=2;C_\mathrm{K};\alpha[s_1,s_2,s_3];T\right]\middle|\,T\right]=\tilde{\mathrm{M}}_P[(c)]+\tilde{\mathrm{M}}_P[(d)]+\tilde{\mathrm{M}}_P[(3rd)]. \qquad \text{--- Eq.(1.3.12)}$$

Here, $\tilde{\mathrm{M}}_P[(c)]$ and $\tilde{\mathrm{M}}_P[(d)]$ are the summed contributions of the type (c)-based and type (d)-based histories, respectively. They are given by exactly the same equations as Eqs.(A1.1.2b,d) and Eqs.(A1.1.2c,e), respectively, with the parameters replaced by those assigned to branch b_1 , and with ΔL^D replaced by ΔL^{D1}. Under Dawg's model, these two terms are given by summations of Eqs.(A1.1.2d',e') with an appropriate replacement of the parameters and the descendant states. Thus, panel A of Figure 3 can also be interpreted as the comparison between the total contribution of these two types of next-fewest-indel histories and the fewest-indel history's contribution. Meanwhile, $\tilde{\mathrm{M}}_P[(3rd)]$ is the contribution from the unique next-fewest-indel history of the 3rd type, Eq.(1.3.11c). According to the definition, Eq.(1.1.2b) supplemented by Eqs.(1.2.4b,6b,8), it is expressed as:

$$\tilde{\mathrm{M}}_P[(3rd)]=\mu_P\left[s^{Root}=s_1,s_0^{Root},n^{Root};C_\mathrm{K}\right]\exp\left\{-\sum_{m=1,2,3}\int_{t_I}^{t_{F:m}}dt\,\delta R_{X:m}^{ID}(s^{Root}=s_1,s_0^{Root},t)\Big|_{b_m}\right\}$$
$$\times\prod_{m=2,3}\int_{t_I}^{t_{F:m}}dt\left[r_{D:m}(2,\Delta L^{D1}+1;s^{Root}=s_1,t)\exp\left\{-\int_t^{t_{F:i}}d\tau\,\delta R_{X:m}^{ID}(s_m,s^{Root}=s_1,t)\right\}\Big|_{b_m}\right]. \qquad \text{--- Eq.(1.3.13)}$$

Under Dawg's model, we have $\delta R_{X:m}^{ID}(s^{Root}=s_1,s_0^{Root},t)\big|_{b_m}=(\lambda_{I:m}+\lambda_{D:m})\Delta L^{D1}$ for $m=1,2,3$, and $\delta R_X^{ID}(s_m,s^{Root}=s_1,t)\big|_{b_m}=-(\lambda_{I:m}+\lambda_{D:m})\Delta L^{D1}$ for $m=2,3$. Moreover, if we assume the uniform distribution of the root sequence length, we have $\mu_P\left[s^{Root}=s_1,s_0^{Root},n^{Root};C_\mathrm{K}\right]=1$. Thus, Eq.(1.3.13) is reduced to:

$$\tilde{\mathrm{M}}_P[(3rd)]=\exp\left(-(\lambda_{I:1}+\lambda_{D:1})\Delta L^{D1}(t_{F:1}-t_I)\right)$$
$$\times\prod_{m=2,3}\left\{\lambda_{D:m}f_{D:m}(\Delta L^{D1})\frac{1-\exp\left(-(\lambda_{I:m}+\lambda_{D:m})\Delta L^{D1}(t_{F:m}-t_I)\right)}{(\lambda_{I:m}+\lambda_{D:m})\Delta L^{D1}}\right\}. \qquad \text{--- Eq.(1.3.13')}$$

Figure 5 shows the ratio of $\tilde{\mathrm{M}}_P[(3rd)]$ to the fewest-indel history's contribution, Eq.(1.3.10), when all three branches have the same length and are assigned the same indel model as that used for Figure 3. Because the ratio compares the multiplication factors concerning the indel events along different branches, its value actually depends on several factors. It would be convenient to keep in mind that the ratio could be approximated by $\lambda_{D:2}f_{D:2}(\Delta L^{D1})(t_{F:2}-t_I)\,\lambda_{D:3}f_{D:3}(\Delta L^{D1})(t_{F:3}-t_I)\big/\lambda_{I:1}f_{I:1}(\Delta L^{D1})(t_{F:1}-t_I)$ when

$(\lambda_{I:m}+\lambda_{D:m})\Delta L^{D1}(t_{F:m}-t_I)$'s are sufficiently smaller than 1 for all $m=1,2,3$. In general, as ΔL^{D1} gets larger, the ratio is expected to decrease, because the relative frequencies of long indels ($f_{I:m}(\Delta L^{D1})$ and $f_{D:m}(\Delta L^{D1})$) are small in general. The ratio is expected to be much smaller than 1 in general. However, it may become quite large when the relative frequency of deletions compared to insertions (*i.e.*, the ratio $\lambda_{D:m}/\lambda_{I:m}$) is considerably larger than 1, or when the lengths of b_2 and b_3 are much larger than that of b_1 (*i.e.*, $t_{F:2}-t_I, t_{F:3}-t_I >> t_{F:1}-t_I$). Such situations are similar to those causing the "Felsenstein zone" regarding a substitution model, where a non-parsimonious substitution history at a site is most likely to occur along a tree (see, *e.g.*, Chapter 9 of Felsenstein 2004). Under the conditions used to draw Figure 5, an indel history of the 3rd type has a probability much smaller than that of the fewest-indel history. The former is less than 5% of the latter even when as much as 0.4 indels per site are expected to occur. This is probably because the history of the 3rd type requires an exquisite spatial coordination of deletions along different branches. And the result implies that the "Felsenstein zone" of indels should generally be quite narrow, consisting of the cases where a node is connected with branches with extremely unequal lengths.

Case (IV) is represented by the external sequence states, $s_1=\left[L,1,...,\Delta L^{D1},R\right]$, $s_2=\left[L,1,...,i,j,...,\Delta L^{D1},R\right]$, and $s_3=\left[L,R\right]$, with $1\le i+1<j\le \Delta L^{D1}+1$ but $(i,j)\ne(0,\Delta L^{D1}+1)$. (Here "$1,...,0$" and "$\Delta L^{D1}+1,...,\Delta L^{D1}$" should be considered to be empty.) In this case, the phylogenetic correctness condition requires the root state to have sites with ancestries $1,...,i$ and $j,...,\Delta L^{D1}$, on top of the PASs. Thus, we have $s_0^{Root}=s_2=\left[L,1,...,i,j,...,\Delta L^{D1},R\right]$. Here, the minimum number of indels is $N_{\min}\left[C_{\mathrm{K}};\alpha[s_1,s_2,s_3];T\right]=2$. And the set of fewest-indel local histories, $\Lambda_{\Psi}^{ID}\left[N_{\mathrm{K}}=2;C_{\mathrm{K}};\alpha[s_1,s_2,s_3];T\right]$, consists of *two* histories. One starts with the root state $s^{Root}=s_0^{Root}\ (=s_2)$, and is represented as:

$$\left\{\vec{\hat{M}}(b_1)=\left[\hat{M}_I(i+1,j-i-1)\right],\vec{\hat{M}}(b_2)=\left[\ \ \right],\vec{\hat{M}}(b_3)=\left[\hat{M}_D(2,\Delta L^{D1}-j+i+2)\right]\right\}.$$

--- Eq.(1.3.14a)

The other starts with the root state $s^{Root}=s_1=\left[L,1,...,\Delta L^{D1},R\right]$, which differs from $s_0^{Root}\ (=s_2)$. It is represented as:

$$\left\{\vec{\hat{M}}(b_1)=\left[\ \ \right],\vec{\hat{M}}(b_2)=\left[\hat{M}_D(i+2,j)\right],\vec{\hat{M}}(b_3)=\left[\hat{M}_D(2,\Delta L^{D1}+1)\right]\right\}.$$ --- Eq.(1.3.14b)

The portion of the multiplication factor contributed from these two fewest-indel local histories, and that contributed from the next-fewest-indel local histories, are calculated in Appendix A2. Table 3 shows the ratio of the total contribution of the next-fewest-indel local histories to that of the fewest-indel local histories for typical configurations in this case. The table indicates that the fewest-indel local histories give a decent approximation of the local MSA probability, as long as the number of descendant sites and the expected number of indels per site are at most moderate.

2. Goodness of approximation by past indel models

In part I (Ezawa, Graur and Landan 2015a), as well as in Section 1 of this series of paper, we established a theoretical formulation that can calculate the probabilities of given alignments (PWAs or MSAs) under a general continuous-time Markov model, which is a *genuine* evolutionary model describing the stochastic sequence evolution along the time axis via insertions and deletions. The Markov model we dealt with can accommodate quite biologically realistic settings, such as overlapping indels, power-law indel length distributions, and indel rate variation across regions. As shown in Subsection 3.1.1 of part I, the probabilities calculated via our theoretical formulation also fulfill, up to a desired order of the perturbation expansion, the Chapman-Kolmogorov (CK) equation, which is a crucial consistency condition for a *genuine* evolutionary model. To the best of our knowledge, the CK equation has never been satisfied by any previous probabilistic theories allowing for indels of multiple residues. Moreover, the results in Section 5 of part I indicated that even the lowest-order approximation based on the fewest-indel histories alone can approximate the probabilities of the local gap configurations of the alignments considerably well, as long as the indel lengths and the branch lengths are at most moderate. Given these ideal properties, the *ab initio* perturbation theory that we formulated in this series of study can be used as a sound "reference point." More precisely, other probabilistic models can be compared to it in order to examine how well they can approximate the biologically realistic probabilities of the alignments, and the conditions under which they can. In this section we will examine the "goodness of approximation" by some representative models of indels in the light of our *ab initio* formulation.

2.1. Geometric indel length distributions

As mentioned in Introduction, a majority of indel probabilistic models that have been used thus far are based on HMMs or transducer theories that calculates the indel component of the probability of an alignment as a product of column-to-column transition probabilities. Such models can only accommodate geometric indel length distributions, or at most linear combinations of two or more geometric distributions. However, empirical analyses revealed that the lengths of inserted/deleted sequences are distributed according to power-laws *e.g.*, Gonnet et al. 1992; Benner et al. 1993; Gu and Li 1995; Kent et al. 2003; Zhang and Gerstein 2003; Chang and Benner 2004; The international Chimpanzee Chromosome 22 Consortium 2004; Yamane et al. 2006; Fan et al. 2007). Therefore, a question naturally arises as to up to what length a geometric distribution can decently approximate a biologically realistic power-law distribution. Here we address this question. As a reference distribution, we used a power-law distribution:

$$f^{PL}(l;a) = l^{-a} \Big/ \sum_{l=1}^{+\infty} l^{-a}, \quad \text{--- Eq.(2.1.1)}$$

with an empirically typical exponent of $a = 1.6$ (panel A of Figure 6). Then we fitted a scaled geometric distribution:

$$f^{SG}(l;A,q) = A \times q^{l-1} \Big/ \sum_{l=1}^{+\infty} q^{l-1} = A(1-q)q^{l-1}, \quad \text{--- Eq.(2.1.2)}$$

to the reference, Eq.(2.1.1), via a least-square fitting across the lengths $l = 1, 2, ..., 100$. We found that the best-fitting parameters are $A = A_{LS}(1.6) \equiv 0.7125$ and $q = q_{LS}(1.6) \equiv 0.3957$. (The distribution is drawn in the dashed line in panel A.) Then, we calculated the ratios:

$$R_{LS}(l;1.6) \equiv f^{SG}(l; A_{LS}(1.6), q_{LS}(1.6)) \big/ f^{PL}(l; a = 1.6), \quad \text{--- Eq.(2.1.3)}$$

for $l = 1, 2, ..., 25$ (panel B of Figure 6). As shown in the figure, $R_{LS}(l; 1.6)$ is less than 1/3 at $l = 5$, less than 1/10 at $l = 7$, and less than 1/1000 at $l = 13$. This indicates that the probabilistic model of indels based on the geometric distribution is somewhat reliable for indels that are at most 4 bases long, which account for 70.5% of all indels if we assume this power-law distribution for all indel lengths. In other words, the geometric distribution substantially underestimates the probabilities of 30% of indels that actually occurred. Although a larger value of q could prevent a rapid damping of the estimated frequency, this then causes the mis-estimation of relative frequencies among commonly occurring indels with, *e.g.*, $l = 1,2,3$ (the dotted line in panel A of Figure 6). Hence, this simple analysis clearly demonstrates how important it is to incorporate biologically realistic indel length distributions into the evolutionary model.

2.2. Models pursuing biological realism

Some existing probabilistic models of indels, such as the "long indel" model (Miklós et al. 2004) and the model of Kim and Sinha (2007), aim to pursue more biological realism by accommodating indel length distributions that are biologically more realistic than geometric distributions. As noted in Subsection 5.1 of part I (and proved in Appendix A6 of part I), as far as each LHS equivalence class is concerned, the probability given by the recipe of Miklós et al. (2004) equals the probability calculated by our formulation under the space- and time-homogeneous model parameters (given in Eqs.(2.4.5a,b) of part I). Thus, regarding the probability of a PWA, the two methods should give the identical sum of contributions from the fewest-indel histories. And, *provided that* both methods can correctly enumerate all possible local histories of any given number of indels consistent with each local gap configuration, they will also give the identical sum of contributions from the histories with up to a desired number of indels. Fulfilling this last condition, however, will be an important task left for future studies. (At present, we cannot tell whether or not the algorithm of Miklós et al. (2004) can correctly enumerate indel histories, because their paper has no explicit description on this topic. So, we will not discuss their method further in this section.)

Here, we specifically examine the model of Kim and Sinha (2007) in the light of our theoretical formulation. Their model is a block-wise HMM, and calculates the probability of a PWA between the ancestral and descendant sequences along a branch as a product of block-wise probabilities. In their HMM, a block is either a column of a PAS, a run of gaps in the ancestor aligned with a run of residues in the descendant, or a run of gaps in the descendant aligned with a run of residues in the ancestor. Each PWA is actually a part of a MSA of given sequences at the external nodes and one of alternative sets of sequences at internal nodes. Ancestral gaps aligned with descendant gaps are removed before evaluating the probability of a PWA. Because their purpose is to find a most likely indel history and a resulting set of consistent ancestral sequence states at internal nodes, they are not interested in an indel event that begins and/or ends in the middle of a block. Thus, they only consider those events that insert/delete the entire blocks in single steps.

We now calculate the probabilities of the local PWAs that were considered in the cases (i)-(iv) in Subsection 1.2, via the model of Kim and Sinha (2007). And we compare the results to those via our theoretical formulation under Dawg's parameters (Eqs.(2.4.4a,b) of part I). In the following, the probabilities via Kim and Sinha (2007) will be calculated according to Eq.(2) and Figure 1C of their paper, and the probabilities via our formulation will be calculated according to the prescriptions in

Subsection 1.2 (and in Appendix A1). We set $t_F - t_I = |b|$ in the following calculations. (Here $|b|$ denotes the length of branch b). Via the model of Kim and Sinha (2007), the PWA probability in case (i) is calculated as:

$$P_{KS}\left[case\,(i)\right] = (1-p_I)^2\,(1-p_D)^2\,, \qquad \text{--- Eq.(2.2.1)}$$

where p_I and p_D are the "transition probabilities of the 'Insertion' and 'Begin deletion' states," respectively. Via our formulation, the PWA probability is:

$$P_{ref}\left[case\,(i)\right] = \exp\left(-\Delta^{Dawg}\,|b| - 2(\lambda_I+\lambda_D)\,|b|\right) \times\left[1+\tilde{\mu}_P^{(2)}[case\,(i)] + \sum_{n=3}^{+\infty}\tilde{\mu}_P^{(n)}[case\,(i)]\right]. \qquad \text{--- Eq.(2.2.2)}$$

Here, Δ^{Dawg} is the abbreviation of the "universal" factor for the indel exit rate (*i.e.*, $\Delta^{Dawg}[\lambda_I, \lambda_D, f_D(.)]$ in Eq.(2.4.4c) of part I). And $\tilde{\mu}_P^{(n)}[case\,(i)]$ is short for $\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = n; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$, *i.e.*, the summed contribution from the n-event local histories to the multiplication factor in case (i). Especially, $\tilde{\mu}_P^{(2)}[case\,(i)]$ is concretely expressed in Eq.(1.2.2a). Now, assuming that $(\lambda_I+\lambda_D)\,|b|$ is sufficiently small, we expand the expression in the square brackets into the power series in $\lambda_I\,|b|$ and $\lambda_D\,|b|$, which will collectively be denoted as $\lambda\,|b|$ when considering the order of magnitude. From Eqs.(1.2.2a,b'), we get $\tilde{\mu}_P^{(2)}[case\,(i)] = \sum_{l=1}^{L^{CO}} \lambda_I f_I(l)\,\lambda_D f_D(l)\,|b|^2/2 + O\left((\lambda\,|b|)^3\right)$. Moreover, the expansion of $\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = n; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$ generally starts with $O\left((\lambda\,|b|)^n\right)$ terms. Thus, we have:

$$P_{ref}\left[case\,(i)\right] = \exp\left(-\Delta^{Dawg}\,|b| - 2(\lambda_I+\lambda_D)\,|b|\right) \times\left[1+\lambda_I\lambda_D\,|b|^2\left(\sum_{l=1}^{L^{CO}} f_I(l)\,f_D(l)\Big/2\right) + O\left((\lambda\,|b|)^3\right)\right]. \qquad \text{--- Eq.(2.2.2')}$$

This and Eq.(2.2.1) will provide the baseline when examining the probabilities in Kim and Sinha's model in other cases.

In case (ii), the PWA probability under the HMM of Kim and Sinha (2007) is calculated as:

$$P_{KS}\left[case\,(ii)\right] = (1-p_I)^3(1-p_D)^2\,p_D\,\mathrm{Pr}_D(\Delta L^A)\,. \qquad \text{--- Eq.(2.2.3)}$$

Here $\mathrm{Pr}_D(l)$ is the "probability distribution on the deletion length (l)," which is assumed as shared among different branches. To facilitate the comparison, we consider the ratio of the probability in case (ii) to that in case (i), which yields:

$$P_{KS}\left[case\,(ii)\right]/P_{KS}\left[case\,(i)\right] = (1-p_I)\,p_D\,\mathrm{Pr}_D(\Delta L^A)\,. \qquad \text{--- Eq.(2.2.4)}$$

Meanwhile, the probability via our formulation is:

$$P_{ref}\left[case\,(ii)\right] = \exp\left(-\Delta^{Dawg}\,|b| - (\lambda_I+\lambda_D)(2+\Delta L^A)\,|b|\right)\times\left[\sum_{n=1}^{+\infty}\tilde{\mu}_P^{(n)}[case\,(ii)]\right]. \qquad \text{--- Eq.(2.2.5)}$$

Here, $\tilde{\mu}_P^{(1)}[case\,(ii)]$ is given by Eq.(1.2.4'), and $\tilde{\mu}_P^{(2)}[case\,(ii)]$ is given by Eq.(1.2.5a) supplemented with Eqs.(1.2.5b,c,d'e'). The ratio of Eq.(2.2.5) to Eq.(2.2.2) is:

$$P_{ref}\left[case\,(ii)\right]/P_{ref}\left[case\,(i)\right] = e^{-(\lambda_I+\lambda_D)\Delta L^A|b|}\left[\sum_{n=1}^{+\infty}\tilde{\mu}_P^{(n)}[case\,(ii)]\right]\left[1+\sum_{n=2}^{+\infty}\tilde{\mu}_P^{(n)}[case\,(i)]\right]^{-1}. \qquad \text{--- Eq.(2.2.6)}$$

Expanding this expression into the power series in $\lambda|b|$, we have:

$$P_{ref}\left[case\,(ii)\right]/P_{ref}\left[case\,(i)\right] = \lambda_D|b|f_D(\Delta L^A) + \tfrac{1}{2}\lambda_D(\lambda_I+\lambda_D)|b|^2 G_D(\Delta L^A) + O\left((\lambda|b|)^3\right). \quad \text{--- Eq.(2.2.6'a)}$$

Here $G_D(\Delta L^A)$ is defined as:

$$G_D(\Delta L^A) \equiv -\Delta L^A f_D(\Delta L^A) + \frac{\lambda_D}{\lambda_I+\lambda_D}\sum_{l=1}^{\Delta L^A-1}(\Delta L^A - l + 1) f_D(l) f_D(\Delta L^A - l) + \frac{\lambda_I}{\lambda_I+\lambda_D}(\Delta L^A+1)\sum_{l=1}^{\min\{L_I^{CO},\,L_D^{CO}-\Delta L^A\}} f_I(l) f_D(\Delta L^A + l) \quad . \quad \text{--- Eq.(2.2.6'b)}$$

Figure 7 shows the ratio $G_D(\Delta L^A)/f_D(\Delta L^A)$ as a function of ΔL^A.

Similarly, according the HMM of Kim and Sinha (2007), the ratio of the PWA probability in case (iii) to that in case (i) is expressed as:

$$P_{KS}\left[case\,(iii)\right]/P_{KS}\left[case\,(i)\right] = \frac{p_I}{1-p_I}\Pr_I(\Delta L^D). \quad \text{--- Eq.(2.2.7)}$$

Here $\Pr_I(l)$ is the "probability distribution on the insertion length (l)," which also is assumed as shared among different branches. The ratio via our formulation is obtained by the power-series expansion in $\lambda|b|$ of Eq.(A1.1.1') and Eq.(A1.1.2a) supplemented with Eqs.(A1.1.2b,c,d',e') (all in Appendix). The result is:

$$P_{ref}\left[case\,(iii)\right]/P_{ref}\left[case\,(i)\right] = \left[\sum_{n=1}^{+\infty}\tilde{\mu}_P^{(n)}[case\,(iii)]\right]\left[1+\sum_{n=2}^{+\infty}\tilde{\mu}_P^{(n)}[case\,(i)]\right]^{-1} = \lambda_I|b|f_I(\Delta L^D) + \tfrac{1}{2}\lambda_I(\lambda_I+\lambda_D)|b|^2 G_I(\Delta L^D) + O\left((\lambda|b|)^3\right). \quad \text{--- Eq.(2.2.8a)}$$

Here $G_I(\Delta L^D)$ is defined as:

$$G_I(\Delta L^D) \equiv -\Delta L^D f_I(\Delta L^D) + \frac{\lambda_I}{\lambda_I+\lambda_D}\sum_{l=1}^{\Delta L^D-1}(\Delta L^D - l + 1) f_I(\Delta L^D - l) f_I(l) + \frac{\lambda_D}{\lambda_I+\lambda_D}(\Delta L^D+1)\sum_{l=1}^{\min\{L_D^{CO},\,L_I^{CO}-\Delta L^D\}} f_I(\Delta L^D + l) f_D(l) \quad . \quad \text{--- Eq.(2.2.8b)}$$

Thanks to the symmetry of the probabilities under the time reversal, Figure 7 also gives the ratio $G_I(\Delta L^D)/f_I(\Delta L^D)$ as a function of ΔL^D, when calculated under the same setting.

Now we compare the results, Eq.(2.2.4) and Eq.(2.2.7), under Kim and Sinha's model with the corresponding results, Eq.(2.2.6'a) and Eq.(2.2.8a), obtained via our formulation. In the method of Kim and Sinha (2007), the substitutions $p_I = c_I|b|$ and $p_D = c_D|b|$ are made first. Then c_I and c_D are estimated from the total frequencies of insertions and deletions, respectively, along the external branches observed from the input MSA. Similarly, $\Pr_I(\Delta L^A)$ and $\Pr_D(\Delta L^A)$ are estimated from the observed length histograms for insertions and deletions, respectively. Let $\{b\}_{PE}$ be the set of branches used for parameter estimations. Then, using the summations of the "reference" results, Eq.(2.2.6'a) and Eq.(2.2.8a), both over $\{b\}_{PE}$, we expect to have:

$$\frac{E\left[c_D\right]}{\lambda_D} = 1 + \frac{\lambda_I+\lambda_D}{2}\langle|b|\rangle_{|b|}\sum_{\Delta L^A=1}^{L_D^{CO}} G_D(\Delta L^A) + O\left((\lambda|b|)^2\right), \quad \text{--- Eq.(2.2.9a)}$$

$$\frac{E[c_I]}{\lambda_I} = 1 + \frac{\lambda_I + \lambda_D}{2} \langle |b| \rangle_{|b|} \sum_{\Delta L^D=1}^{L_T^{CO}} G_I(\Delta L^D) + O\left((\lambda |b|)^2\right), \quad \text{--- Eq.(2.2.9b)}$$

$$\frac{E[c_D]}{\lambda_D} E\left[\Pr_D(\Delta L^A)\right] = f_D(\Delta L^A) + \frac{\lambda_I + \lambda_D}{2} \langle |b| \rangle_{|b|} G_D(\Delta L^A) + O\left((\lambda |b|)^2\right),$$

--- Eq.(2.2.9c)

$$\frac{E[c_I]}{\lambda_I} E\left[\Pr_I(\Delta L^D)\right] = f_I(\Delta L^D) + \frac{\lambda_I + \lambda_D}{2} \langle |b| \rangle_{|b|} G_I(\Delta L^D) + O\left((\lambda |b|)^2\right).$$

--- Eq.(2.2.9d)

Here $E[X]$ denotes the expected value of the estimated parameter X, which is the average of estimated X over all indel processes under the given set of conditions (the tree and model parameters). And we also used the notation, $\langle X(b) \rangle_{|b|} \equiv \left[\sum_{b \in \{b\}_{PE}} |b| X(b)\right] \Big/ \sum_{b \in \{b\}_{PE}} |b|$. [NOTE: The actual values of c_I and c_D estimated by the method of Kim and Sinha (2007) may be slightly smaller than Eqs.(2.2.9b,c), because the denominator in their method is the total number of MSA columns, instead of the average numbers of possible indel positions in ancestral sequences.] Usually, $\frac{1}{2}(\lambda_I + \lambda_D)\langle |b| \rangle_{|b|}$ is quite small, at most $O\left(10^{-1}\right)$ and typically $O\left(10^{-2}\right)$. Thus, as long as the actual parameters, λ_I, λ_D, $f_I(\Delta L^D)$, and $f_D(\Delta L^A)$, do not considerably vary across branches, and provided that the MSA is sufficiently long and accurate, the estimated values of c_I and c_D, respectively, should approximate λ_I and λ_D fairly well. Also, under the same situation, the estimated values of $\Pr_I(l)$ and $\Pr_D(l)$, respectively, should approximate $f_I(l)$ and $f_D(l)$ fairly well, as long as the ratios $G_I(l)/f_I(l)$ and $G_D(l)/f_D(l)$ are sufficiently less than $\left[\frac{1}{2}(\lambda_I + \lambda_D)\langle |b| \rangle_{|b|}\right]^{-1}$. However, it does not actually matter so much whether or not the estimated Kim-Sinha parameters (c_I, c_D, $\Pr_I(l)$ and $\Pr_D(l)$) approximate Dawg's indel parameters (λ_I, λ_D, $f_I(l)$ and $f_D(l)$) fairly well. What actually matters is how accurately Eq.(2.2.4) and Eq.(2.2.7) approximate Eq.(2.2.6'a) and Eq.(2.2.8a), respectively, using the estimated parameters. In an extreme case where all branches have the same branch length, the approximation should be nearly perfect. This is because, in this case, $|b| \approx \langle |b| \rangle_{|b|}$ for all branches, and thus because we can use Eqs.(2.2.9a-d) without any significant modifications to estimate the probabilities (not involving case (iv)). [It should be noted here that Eq.(2.2.4) and Eq.(2.2.7), respectively, contain extra multiplication factors, $(1 - p_I)$ and $(1 - p_I)^{-1}$, compared to the corresponding Eq.(2.2.6'a) and Eq.(2.2.8a). However, these factors should remain close to 1, because $p_I = c_I |b|$ should normally be at most $O\left(10^{-1}\right)$.] In contrast, the approximations by Eq.(2.2.4) and Eq.(2.2.7) could considerably deteriorate, *e.g.*, when $\left|(\lambda_I + \lambda_D)(|b| - \langle |b| \rangle_{|b|})\right|$ times the ratios, $G_D(\Delta L^A)/f_D(\Delta L^A)$ and $G_I(\Delta L^D)/f_I(\Delta L^D)$, respectively, become comparable to or greater than 1 (unity). Because $G_D(\Delta L^A)$ and $G_I(\Delta L^D)$ are mostly contributed from next-fewest-indel local histories containing overlapping indels, we can interpret the result as follows. "Overlapping indels start to make Kim and Sinha's method poorly approximate the alignment probabilities when the involved gap is long and the branch lengths show a large variation." If there is a

good reason to believe that the Dawg indel parameters (λ_I, λ_D, $f_I(l)$ and $f_D(l)$) are shared among all branches, one way to mitigate the aforementioned effects of overlapping indels may be to set:

$$c_D(|b|)\Pr_D(l,|b|) = \lambda_D f_D(l) + \lambda_D \frac{\lambda_I + \lambda_D}{2}|b|G_D(l), \quad \text{--- Eq.(2.2.10a)}$$

$$c_I(|b|)\Pr_I(l,|b|) = \lambda_I f_I(l) + \lambda_I \frac{\lambda_I + \lambda_D}{2}|b|G_I(l), \quad \text{--- Eq.(2.2.10b)}$$

and to fit λ_I, λ_D, $f_I(l)$ and $f_D(l)$ according to these equations supplemented with Eqs.(2.2.6'b,8b). Now, as indicated by Figure 7, under the power-law indel length distributions, the ratios $G_D(\Delta L^A)/f_D(\Delta L^A)$ and $G_I(\Delta L^D)/f_I(\Delta L^D)$ are less than 4 in absolute value when the gap is 300 residues long or shorter. Therefore, the 2nd-order terms will begin to substantially influence the results when $\left|\frac{1}{2}(\lambda_I + \lambda_D)(|b| - \langle |b| \rangle_{|b|})\right|$ is larger than, say, 0.1. Such a situation will be quite rare in practical sequence analyses. Even if we encounter such a rare case, then local histories with more than 2 indels will begin to account for a substantial fraction of the probability. Considering this way, we expect that the method of Kim and Sinha (2007) will pretty well approximate the probabilities of local PWAs belonging to cases (ii) and (iii) at least within the threshold $(\Delta L^A)_{0.5}^{(2-ii)}$ defined in Subsection 1.2.

Finally, we consider case (iv). Indel histories giving rise to the local sequence states in this category are shown, *e.g.*, in Figure 5, panels F and G of Figure 3, and panel A of Figure 6 (all of them are in part I). In such a situation, an aligner will reconstruct a MSA that is like either panel B or C of Figure 5 of part I (if the reconstruction is correct), and Kim-Sinha's method assigns a probability according to the reconstructed MSA. Whether it is like panel B or panel C, the assigned probability is the same, and its ratio to the probability of case (i) is:

$$P_{KS}[case\ (iv)]/P_{KS}[case\ (i)] = p_I \Pr_I(\Delta L^D)\, p_D \Pr_D(\Delta L^A). \quad \text{--- Eq.(2.2.11)}$$

Via our formulation, how to calculate the probability in this case was briefly described in the middle of Subsection 1.2, and detailed in Appendix A1.2. In this case, each fewest-indel history consists of two indels, and each next-fewest-indel history consists of three indels. Because there are as many as 24 types of next-fewest-indel histories, here we only consider the fewest-indel histories. Then, the lowest-order contribution of the multiplication factor, $\tilde{\mu}_P^{(2)}[case\ (iv)]$, is given by Eq.(A1.2.1a), supplemented with Eqs.(A1.2.1b,c,d,e',f',g'). Expanding each term into a power series in $\lambda|b|$, we get the following expression for the ratio:

$$\begin{aligned} &P_{ref}[case\ (iv)]/P_{ref}[case\ (i)] \\ &= e^{-(\lambda_I+\lambda_D)\Delta L^A|b|}\left[\sum_{n=2}^{+\infty}\tilde{\mu}_P^{(n)}[case\ (iv)]\right]\left[1+\sum_{n=2}^{+\infty}\tilde{\mu}_P^{(n)}[case\ (i)]\right]^{-1} \\ &= \lambda_D\lambda_I|b|^2\left[\tfrac{1}{2}f_D(\Delta L^A)f_I(\Delta L^D) + \left\{\sum_{l=0}^{\min\{L_I^{CO}-\Delta L^D,\, L_D^{CO}-\Delta L^A\}} f_I(\Delta L^D + l)f_D(\Delta L^A + l)\right\}\right] \\ &\qquad + O\left((\lambda|b|)^3\right). \end{aligned}$$

--- Eq.(2.2.12)

In case (iv), as opposed to in cases (ii) and (iii), the PWA probabilities via the HMM of Kim and Sinha (2007) differ considerably from that via our formulation *even when* $\langle |b| \rangle_{|b|} << 1$ and $|b| << 1$. Under these conditions, $p_I \Pr_I(\Delta L^D)$ and $p_D \Pr_D(\Delta L^A)$ quite

accurately approximate $\lambda_I |b| f_I(\Delta L^D)$ and $\lambda_D |b| f_D(\Delta L^A)$, respectively (see Eqs.(1.2.9c,d)), and the $O\left((\lambda |b|)^3\right)$ terms in Eq.(2.2.12) can be neglected. Thus, we have:

$$\frac{P_{ref}\left[case\ (iv)\right]/P_{ref}\left[case\ (i)\right]}{P_{KS}\left[case\ (iv)\right]/P_{KS}\left[case\ (i)\right]} \approx \frac{3}{2} + \sum_{l=1}^{\min\{L_I^{CO}-\Delta L^D,\, L_D^{CO}-\Delta L^A\}} \frac{f_I(\Delta L^D + l) f_D(\Delta L^A + l)}{f_D(\Delta L^A) f_I(\Delta L^D)} .$$

--- Eq.(2.2.13)

Table 4 shows this ratio for typical cases. The second term on the right hand side of Eq.(2.2.13) is the effect of overlapping indels. When $\Delta L^A = \Delta L^D = 1$, this term is expected to be quite small; for example, it is about 0.167 if $f_I(l) = f_D(l) \propto l^{-1.6}$. And it gets more and more influential when ΔL^A and/or ΔL^D gets larger, and it substantially exceeds 1 (unity) in some cases (Table 4). Actually, a similar effect was incorporated in the HMM of Knudsen and Miyamoto (2003). Their HMM could only accommodate geometric indel length distributions, and consequently the relevant term was independent of ΔL^A and ΔL^D. Coming back to Eq.(2.2.13), the first term on the right hand side, 3/2, differs from 1 (unity) because the HMM of Kim and Sinha (2007) does not correctly take account of the non-overlapping indel histories, either. This error is actually shared by most of the standard, or nearly standard, HMMs and transducers used thus far as probabilistic models of indels (see Background of part I). Taking these results into consideration, another possible improvement on the model of Kim and Sinha (2007) would be to modify the HMM structure so that the probability of an insertion and an immediately adjacent deletion (or that of the opposite configuration) will be given by Eq.(2.2.12), or by its extension to include the terms of higher-orders in $\lambda |b|$.

2.3. Violation of Chapman-Kolmogorov equation

The Chapman-Kolmogorov (CK) equation (Eq.(3.1.1.1) in part I),

$$\hat{P}^{ID}(t_I, t_M)\, \hat{P}^{ID}(t_M, t_F) = \hat{P}^{ID}(t_I, t_F) \quad (t_I < t_M < t_F),$$

is a crucial condition that must be satisfied by *genuine* evolutionary models based on any continuous-time Markov models. Thus far, the CK equation was violated by most HMMs and transducer theories, except those directly derived from evolutionary models, *e.g.*, the TKF91 model (Thorne et al. 1991). As formally shown in Appendix A3 of part I (Ezawa, Graur and Landan 2015a), our theoretical formulation could satisfy the CK equation up to any desired order of the perturbation expansion. Therefore, our formulation enables us to analytically examine the effects of the violation of the CK equation. Here, we will only consider a simplest yet non-trivial example, *i.e.*, case (ii) in Subsection 1.2. And we will examine the HMM of Kim and Sinha (2007) because of its generality; most HMMs could be obtained from their HMM by slightly modifying the parameters and by fixing the indel length distributions to geometric ones, possibly with time-dependent gap-extension probabilities.

Via the HMM of Kim and Sinha (2007), the PWA probability in case (ii) was given by Eq.(VII-2.3), with the substitutions $p_I = c_I |b|$ and $p_D = c_D |b|$:

$$P_{KS}\left[case\ (ii)\right] = (1 - c_I |b|)^3 (1 - c_D |b|)^2\, c_D |b| \Pr_D(\Delta L^A) . \text{ --- Eq.(2.3.1)}$$

In contrast, the probability in the same case via our formulation was given by Eq.(2.2.5), which can be partially expanded into the power-series in $\lambda |b|$ as:

$$P_{ref}\left[case\ (ii)\right] = \exp\left\{-\Delta^{Dawg}|b| - 2(\lambda_I+\lambda_D)|b|\right\}$$
$$\times\left[\lambda_D|b|f_D(\Delta L^A) + \tfrac{1}{2}\lambda_D(\lambda_I+\lambda_D)|b|^2 G_D(\Delta L^A) + O\left((\lambda|b|)^3\right)\right]. \quad \text{--- Eq.(2.3.2)}$$

Here $G_D(\Delta L^A)$ was given in Eq.(2.2.6'b). And $\exp\left\{-\Delta^{Dawg}|b|\right\}$ is a "universal factor" that occur in the probability of any indel processes on the entire sequence (along branch b). Eq.(2.3.2) satisfies the CK equation up to (and including) $O\left((\lambda|b|)^2\right)$. Now, compare Eq.(2.3.1) with Eq.(2.3.2), aside from the universal factor. First, if we set $c_I = \lambda_I$ and $c_D = \lambda_D$, $(1-c_I|b|)^2(1-c_D|b|)^2$ should approximate $\exp\left\{-2(\lambda_I+\lambda_D)|b|\right\}$ quite well, because $\lambda_I|b|$ and $\lambda_D|b|$ should usually be at most $O\left(10^{-1}\right)$. Thus, it is sufficient to compare $(1-c_I|b|)c_D|b|\Pr_D(\Delta L^A)$ with the expression in the square brackets on the right hand side of Eq.(2.3.2). If $\Pr_D(\Delta L^A)$ is considered as independent of $|b|$, it should be equal to $f_D(\Delta L^A)$. Then, whether or not Eq.(2.3.1) approximately satisfies the CK equation depends on whether or not the absolute value of the difference, $\left|\left\{\tfrac{1}{2}\lambda_D(\lambda_I+\lambda_D)|b|^2 G_D(\Delta L^A)\right\} - \left\{-c_I c_D|b|^2 \Pr_D(\Delta L^A)\right\}\right|$, is negligible compared to $\lambda_D|b|f_D(\Delta L^A)$. As Figure 7 indicates, this condition will hold if ΔL^A is within a "critical value," which decreases as $(\lambda_I+\lambda_D)|b|$ increases. Once ΔL^A exceeds the critical value, the violation of the CK equation will considerably impact on the probability estimation, and therefore on the data analyses. Under the parameter setting used for Figure 7, the critical value is $\Delta L^A_{cr} \approx 300$ when $(\lambda_I+\lambda_D)|b| \approx 0.6$. This means that, *as far as case (ii) local gap-configurations are concerned*, Kim and Sinha's HMM practically satisfies the CK equation at least up to $O\left((\lambda|b|)^2\right)$. However, we expect that their HMM will severely violate the CK equation when it is applied to case (iii) local PWAs, in view of the results in Subsection 2.2.

Some past simulation analyses (*e.g.*, Thorne et al. 1992; Knudsen and Miyamoto 2003; Metzler 2003) seem to have concluded that the violation of the CK equation, or its cause, *i.e.*, the failure to accommodate overlapping indels, did not seriously impact on the results of data analyses. These results might be because they used geometric distributions of indel lengths. To see if this is indeed the case, let's assume $f_I(l) = f_D(l) = (1-q)q^{l-1}$. Substituting them into Eq.(2.2.6'b), we have:

$$G_D(\Delta L^A) \approx f_D(\Delta L^A)\times\left[-\Delta L^A + \frac{\lambda_D}{\lambda_I+\lambda_D}\frac{(\Delta L^A-1)(\Delta L^A+2)(1-q)}{2q} + \frac{\lambda_I}{\lambda_I+\lambda_D}\frac{(\Delta L^A+1)q}{1+q}\right]. \quad \text{--- Eq.(2.3.3)}$$

Here we used $q^{2\min\{L_I^{CO}, L_D^{CO}-\Delta L^A\}} << 1$ to obtain the approximation. Now, with the approximations $c_I \approx \lambda_I$, $c_D \approx \lambda_D$ and $\Pr_D(\Delta L^A) \approx f_D(\Delta L^A)$, we calculate the ratio:

$$\left[\left\{\tfrac{1}{2}\lambda_D(\lambda_I+\lambda_D)|b|^2 G_D(\Delta L^A)\right\} - \left\{-c_I c_D|b|^2 \Pr_D(\Delta L^A)\right\}\right] \Big/ \left[\lambda_D|b|f_D(\Delta L^A)\right]$$
$$\approx -(\lambda_I+\lambda_D)|b|\frac{\Delta L^A}{2} + \lambda_D|b|\frac{(\Delta L^A-1)(\Delta L^A+2)(1-q)}{4q} + \lambda_I|b|\left(\frac{(\Delta L^A+1)q}{2(1+q)}+1\right). \quad \text{--- Eq.(2.3.6)}$$

If ΔL^A is quite large, *e.g.*, 10 or greater, the middle term on the right hand side will predominate, and the ratio could be roughly approximated as

$\lambda_D |b| (\Delta L^A)^2 (1-q)/(4q)$. Thus, the "critical value" is roughly given by $\Delta L^A_{cr} \approx \sqrt{4q/(\lambda_D |b| (1-q))}$. If we set, *e.g.*, $\lambda_D = 0.1$ and $q = q_{LS}(1.6) = 0.3957$ as given above Eq.(2.1.3), the critical value would be: $\Delta L^A_{cr} \approx 8$ for $|b| = 0.4$, and $\Delta L^A_{cr} \approx 26$ for $|b| = 0.04$. At these critical values, geometric deletion length frequencies are already pretty small; we have $f_D(\Delta L^A = 8) \approx 9.2 \times 10^{-4}$ and $f_D(\Delta L^A = 26) \approx 5.2 \times 10^{-11}$ when $q = q_{LS}(1.6) = 0.3957$. Therefore, as long as the geometric indel length distributions are used for simulations, it would be very rare to encounter indels whose lengths exceed the critical value. Moreover, as shown in Subsection 2.1, when the indels are longer than these critical values, models with geometric indel length distributions severely underestimate their frequencies, to the extent that they are deemed almost useless. With these circumstances compounded together, it would come as no surprise that the past studies did not detect significant impacts of the violation of the CK equation on their results. [NOTE: Precisely speaking, (nearly) standard HMMs and transducers can take account of *a part* of contributions from multiple deletions. For example, the total contribution of two-deletion histories that they can capture is $(\Delta L^A - 1)(\lambda_D |b|)^2 (1-q) f_D(\Delta L^A)$. Here $(\Delta L^A - 1)$ comes from the number of ways in which a run of gaps can be split into two segments. Thus, its ratio to the first-order term, $\lambda_D |b| f_D(\Delta L^A)$, is $(\Delta L^A - 1)(\lambda_D |b|)(1-q)$. This is roughly $4q/\Delta L^A$ times as large as the dominant term in Eq.(2.3.6). This ratio is much smaller than 1 (unity) when ΔL^A is around or greater than the above critical values. Therefore, such histories with two contiguous deletions have only negligible effects on the present argument.]

In contrast, if we incorporate biologically realistic indel length distributions into indel probabilistic models that are *not genuine* evolutionary models, such effects may become remarkable when dealing with long indels and/or long branches. The effects, however, might be somewhat alleviated by fitting indel rate parameters and length distributions that depend on the branch length as in Eqs.(2.2.10a,b).

Conclusions

In a previous paper (Ezawa, Graur and Landan 2015a), we approached the fundamental problems on the probability of indels in an absolutely orthodox manner. More specifically, we established the theoretical basis of an *ab initio* perturbative formulation of a general continuous-time Markov model, which is a *genuine* stochastic model describing the evolution of an *entire* sequence via indels along the time axis. Using the formulation, we proved that, if the indel model parameters satisfy a certain set of conditions, the probability of an alignment is indeed factorable into the product of an overall factor and contributions from local alignments delimited by preserved ancestral sites (PASs).

In this paper, we concretely calculated the probability contributions from individual local alignments using perturbation analyses. The results indicated that even the fewest-indel terms alone can approximate the probabilities pretty well as long as the branch lengths and the indel lengths are at most moderate. And we also clarified the parameter regions where the alignment probabilities are safely approximated by the hidden Markov models (HMMs) of indels (and by association the transducer theories) that were often used in the past. We showed that the approximation by the geometric indel length distributions grossly deteriorates as soon as the indels become longer than 4 sites. We also showed that these HMMs violate the Chapman-Kolmogorov (CK) equation because they lack most of the non-fewest-indel terms in the perturbation expansion. This finding enables to predict the regions where the violation of the CK equation starts to compromise the reliability of these models. The analyses also suggested possible modifications to these models in order to improve their accuracy.

To summarize, by depending purely on the first principle, our *ab initio* perturbation formulation provides a sound reference point to which other indel models can be compared in order to see when and how well they can approximate the true alignment probabilities.

Methods

Implementation of the formulas

Most of the formulas used for the perturbation analyses were implemented in Perl. They are available as a part of the package named LOLIPOG (log-likelihood for the pattern of gaps in MSA), which in turn is available at the FTP repository of the Bioinformatics Organization (Ezawa 2013).

Authors' contributions

KE conceived of and mathematically formulated the theoretical framework in this paper, implemented the key algorithms, participated in designing the study, performed all the mathematical analyses, and drafted the manuscript. DG and GL participated in designing the study, helped with the interpretation of the data, and helped with the drafting of the manuscript.

Acknowledgements

This study is dedicated to the late Dr. Keiji Kikkawa, who was a renowned theoretical physicist, one of the key pioneers of the string field theory of the elementary particle physics, and the best ever mentor of K.E. We are grateful to Dr. R. A. Cartwright at Arizona State University for his useful information and discussions that inspired this study. We appreciate the logistic support and the feedback of Dr. Tetsushi Yada at the Kyushu Institute of Technology. We would also like to thank the three anonymous referees of the predecessor manuscript entitled: "Framework that enables approximate lilelihood analysis of insertions/deletions on multiple sequence alignment." Their comments helped drastically improve the study itself, not to mention the manuscript. This work was a part of the project, "Error Correction in Multiple Sequence Alignments," which was funded by US National Library of Medicine [grant number LM010009-01 to Dan Graur and Giddy Landan at the University of Houston]. The later stage of this work was also supported by Grants-in-Aid No. 221S0002, which was awarded to Tetsushi Yada by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Appendix

A1. Perturbation calculation of multiplication factors for PWAs between ancestral and descendant sequences

A1.1. For case (iii) local PWAs

Here we detail the calculation of the sum of contributions from the fewest-indel events as well as that from the next-fewest-indel events in case (iii) considered in Section 1.2 of Results, that is, the case where the ancestor has no site but the descendant has one or more sites in between a pair of preserved ancestral sites (PASs) (Figure 2 C).

In case (iii), we assume that the descendant state has ΔL^D sites in between the flanking PASs. Thus, the ancestral and descendant states could be represented as $s^A = [L, R]$ and $s^D = [L, \upsilon_1^D, ..., \upsilon_{\Delta L^D}^D, R]$, respectively. The ancestries satisfy $\upsilon_i^D \neq \upsilon_j^D$ for $i \neq j$, and their details depend on the responsible indel history. As long as $\Delta L^D \leq L_I^{CO}$, $N_{\min}[\gamma_\kappa; \alpha(s^A, s^D)] = 1$, and there is only one fewest-indel history, $[\hat{M}_I(1, \Delta L^D)]$, which consists of a single event that inserts the descendant sites in between the PASs. Therefore, the sum of contributions from the fewest-indel history is:

$$\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = 1; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$
$$= \int_{t_I}^{t_F} dt\, r_I(1, \Delta L^D; s^A, t) \exp\left\{-\int_{t_I}^{t} d\tau\, \delta R_X^{ID}(s^A, s^A, t) - \int_{t}^{t_F} d\tau\, \delta R_X^{ID}(s^D, s^A, t)\right\}$$
$$= \int_{t_I}^{t_F} dt\, r_I(1, \Delta L^D; s^A, t) \exp\left\{-\int_{t}^{t_F} d\tau\, \delta R_X^{ID}(s^D, s^A, t)\right\}.$$

--- Eq.(A1.1.1)

As in case (ii), each next-fewest indel history is composed of two indel events, and classified into two types: (c) two successive insertions, $\left[\hat{M}_I(1, \Delta L^D - l), \hat{M}_I(x, l)\right]$ with $l = 1, ..., \Delta L^D - 1$ and $x = 1, ..., \Delta L^D - l + 1$; and (d) an insertion followed by a deletion, $\left[\hat{M}_I(1, \Delta L^D + l), \hat{M}_D(x, x + l - 1)\right]$ with $l = 1, ..., \min\{L_D^{CO}, L_I^{CO} - \Delta L^D\}$ and $x = 2, ..., \Delta L^D + 2$. Thus, in this case, the portion contributed by the next-fewest indel histories is given by:

$$\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = 2; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right] = \tilde{\mu}_P[(c)] + \tilde{\mu}_P[(d)]. \quad \text{--- Eq.(A1.1.2a)}$$

Here,

$$\tilde{\mu}_P[(c)] \equiv \sum_{l=1}^{\Delta L^D - 1} \sum_{x=1}^{\Delta L^D - l + 1} \mu_P\left[\left(\left[\hat{M}_I(1, \Delta L^D - l), \hat{M}_I(x, l)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$

--- Eq.(A1.1.2b)

is the sum of contributions from the histories of type (c), and

$$\tilde{\mu}_P[(d)] \equiv \sum_{l=1}^{\min\{L_D^{CO}, L_I^{CO} - \Delta L^D\}} \sum_{x=2}^{\Delta L^D + 2} \mu_P\left[\left(\left[\hat{M}_I(1, \Delta L^D + l), \hat{M}_D(x, x + l - 1)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$

--- Eq.(A1.1.2c)

is the sum of contributions from the histories of type (d). As in case (i), let $\left\langle s[\Delta L^D - l]\right| \equiv \left\langle s^A \middle| \hat{M}_I(1, \Delta L^D - l)\right.$ be the intermediate state in each type (c) history. Then, the history's contribution is calculated as:

$$\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D - l), \hat{M}_I(x,l)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$

$$= \int_{t_I}^{t_F} dt \int_t^{t_F} dt' \begin{bmatrix} r_I(1,\Delta L^D - l; s^A, t)\, r_I(x,l; s\cdot[\Delta L^D - l], t') \\ \times \exp\left\{-\int_t^{t'} d\tau\, \delta R_X^{ID}(s[\Delta L^D - l], s^A, \tau) - \int_{t'}^{t_F} d\tau\, \delta R_X^{ID}(s^D, s^A, \tau)\right\} \end{bmatrix}.$$

--- Eq.(A1.1.2d)

Similarly, each type (d) history's contribution is calculated as:

$$\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D + l), \hat{M}_D(x, x+l-1)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$

$$= \int_{t_I}^{t_F} dt \int_t^{t_F} dt' \begin{bmatrix} r_I(1,\Delta L^D + l; s^A, t)\, r_D(x, x+l-1; s\cdot[\Delta L^D + l], t') \\ \times \exp\left\{-\int_t^{t'} d\tau\, \delta R_X^{ID}(s[\Delta L^D + l], s^A, \tau) - \int_{t'}^{t_F} d\tau\, \delta R_X^{ID}(s^D, s^A, \tau)\right\} \end{bmatrix}.$$

--- Eq.(A1.1.2e)

Let us now calculate Eqs.(A1.1.1,2a-e) under Dawg's indel model or the "long indel" model, as in the previous cases in Section 1.2 of Results. In this model, we have $\delta R_X^{ID}(s^D, s^A, \tau) = +(\lambda_I + \lambda_D)\Delta L^D$, and Eq.(A1.1.1) becomes:

$$\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = 1; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$

$$= \lambda_I f_I(\Delta L^D) \frac{1 - \exp\left(-(\lambda_I + \lambda_D)\Delta L^D (t_F - t_I)\right)}{(\lambda_I + \lambda_D)\Delta L^D}.$$

--- Eq.(A1.1.1')

Similarly, using $\delta R_X^{ID}(s[\Delta L^D - l], s^A, \tau) = +(\lambda_I + \lambda_D)(\Delta L^D - l)$ and $\delta R_X^{ID}(s[\Delta L^D + l], s^A, \tau) = +(\lambda_I + \lambda_D)(\Delta L^D + l)$, Eqs.(A1.1.2d,e) are calculated, respectively, as:

$$\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D - l), \hat{M}_I(x,l)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$

$$= \frac{\lambda_I f_I(\Delta L^D - l)\lambda_I f_I(l)}{(\lambda_I + \lambda_D) l} \left[\frac{1 - e^{-(\lambda_I + \lambda_D)(\Delta L^D - l)(t_F - t_I)}}{(\lambda_I + \lambda_D)(\Delta L^D - l)} - \frac{1 - e^{-(\lambda_I + \lambda_D)\Delta L^D (t_F - t_I)}}{(\lambda_I + \lambda_D)\Delta L^D}\right],$$

--- Eq.(A1.1.2d')

$$\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D + l), \hat{M}_D(x, x+l-1)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$

$$= \frac{\lambda_I f_I(\Delta L^D + l)\lambda_D f_D(l)}{(\lambda_I + \lambda_D) l} \left[\frac{1 - e^{-(\lambda_I + \lambda_D)\Delta L^D (t_F - t_I)}}{(\lambda_I + \lambda_D)\Delta L^D} - \frac{1 - e^{-(\lambda_I + \lambda_D)(\Delta L^D + l)(t_F - t_I)}}{(\lambda_I + \lambda_D)(\Delta L^D + l)}\right].$$

--- Eq.(A1.1.2e')

Substituting Eqs.(A1.1.2d',e') into Eqs.(A1.1.2a,b,c), we can concretely calculate the total contribution from next-fewest-indels.

A1.2. For case (iv) local PWAs

Here we detail the calculation of the sum of contributions from the fewest-indel events as well as that from the next-fewest-indel events in case (iv) considered in Section 1.2 of Results, that is, the case where both the ancestor and the descendant

have one or more site(s) in between a pair of PASs, but no ancestral site is related to any of the descendant sites (Figure 2 D).

In case (iv), we assume that the ancestral and the descendant states have ΔL^A and ΔL^D sites, respectively, in between the flanking PASs. Thus, the ancestral and descendant states could be represented as $s^A = \left[L, 1, ..., \Delta L^A, R\right]$ and $s^D = \left[L, \upsilon_1^D, ..., \upsilon_{\Delta L^D}^D, R\right]$, respectively. Here, the descendant ancestries satisfy $\upsilon_i^D \notin \{L, 1, ..., \Delta L^A, R\}$ for all $i = 1, ..., \Delta L^D$ and $\upsilon_i^D \neq \upsilon_j^D$ for all $i \neq j$, and their details depend on the responsible indel history. As long as $\Delta L^A \leq L_D^{CO}$ and $\Delta L^D \leq L_I^{CO}$, $N_{\min}\left[\gamma_\kappa; \alpha(s^A, s^D)\right] = 2$. As indicated by Eqs.(A1.3c',d') in Appendix A1 of part I (Ezawa, Graur and Landan 2015a), there are three types of fewest-indel histories: (e) the deletion of the ancestral sites followed by an insertion of ΔL^D sites, $\left[\hat{M}_D(2, \Delta L^A + 1), \hat{M}_I(1, \Delta L^D)\right]$; (f) an insertion immediately on the right of the ancestral sites to be deleted, followed by the deletion, $\left[\hat{M}_I(\Delta L^A + 1, \Delta L^D + l), \hat{M}_D(2, \Delta L^A + l + 1)\right]$ with $l = 0, ..., \min\{L_I^{CO} - \Delta L^D, L_D^{CO} - \Delta L^A\}$; and (g) an insertion immediately on the left of the ancestral sites to be deleted, followed by the deletion, $\left[\hat{M}_I(1, \Delta L^D + l), \hat{M}_D(\Delta L^D + 2, \Delta L^A + \Delta L^D + l + 1)\right]$ also with $l = 0, ..., \min\{L_I^{CO} - \Delta L^D, L_D^{CO} - \Delta L^A\}$. Thus, the sum of contributions by the fewest-indel histories is expressed as:

$$\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = 2; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right] = \tilde{\mu}_P[(e)] + \tilde{\mu}_P[(f)] + \tilde{\mu}_P[(g)]. \quad \text{--- Eq.(A1.2.1a)}$$

Here,

$$\tilde{\mu}_P[(e)] \equiv \mu_P\left[\left(\left[\hat{M}_D(2, \Delta L^A + 1), \hat{M}_I(1, \Delta L^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right] \quad \text{--- Eq.(A1.2.1b)}$$

is the contribution from the single history of type (e),

$$\tilde{\mu}_P[(f)] \equiv \sum_{l=0}^{\min\{L_I^{CO} - \Delta L^D, L_D^{CO} - \Delta L^A\}} \mu_P\left[\left(\left[\hat{M}_I(\Delta L^A + 1, \Delta L^D + l), \hat{M}_D(2, \Delta L^A + l + 1)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$

--- Eq.(A1.2.1c)

is the sum of contributions from the type (f) histories, and

$$\tilde{\mu}_P[(g)] \equiv \sum_{l=0}^{\min\{L_I^{CO} - \Delta L^D, L_D^{CO} - \Delta L^A\}} \mu_P\left[\left(\left[\hat{M}_I(1, \Delta L^D + l), \hat{M}_D(\Delta L^D + 2, \Delta L^A + \Delta L^D + l + 1)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$

--- Eq.(A1.2.1d)

is the sum of contributions from the type (g) histories. To calculate the contribution from each of these histories, let $\langle s_0 | \equiv \langle [L, R] | = \langle s^A | \hat{M}_D(2, \Delta L^A + 1)$ be the intermediate state of the type (e) history. And, as in case (ii), let $\langle s^A \cdot [x, +l] | \equiv \langle s^A | \hat{M}_I(x, l)$ be the state type including the intermediate states of the histories of types (f) and (g). Then, the type (e) history's contribution is:

$$\mu_P\left[\left(\left[\hat{M}_D(2,\Delta L^A+1),\hat{M}_I(1,\Delta L^D)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$

$$=\int_{t_I}^{t_F}dt\int_t^{t_F}dt'\begin{bmatrix} r_D(2,\Delta L^A+1;s^A,t)\,r_I(1,\Delta L^D;s_0,t') \\ \times\exp\left\{-\int_t^{t'}d\tau\,\delta R_X^{ID}(s_0,s^A,\tau)-\int_{t'}^{t_F}d\tau\,\delta R_X^{ID}(s^D,s^A,\tau)\right\}\end{bmatrix},$$

--- Eq.(A1.2.1e)

the contribution from each history of type (f) is:

$$\mu_P\left[\left(\left[\hat{M}_I(\Delta L^A+1,\Delta L^D+l),\hat{M}_D(2,\Delta L^A+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$

$$=\int_{t_I}^{t_F}dt\int_t^{t_F}dt'\begin{bmatrix} r_I(\Delta L^A+1,\Delta L^D+l;s^A,t)\,r_D(2,\Delta L^A+l+1;s^A\cdot[\Delta L^A+1,+\Delta L^D+l],t') \\ \times\exp\left\{-\int_t^{t'}d\tau\,\delta R_X^{ID}(s^A\cdot[\Delta L^A+1,+\Delta L^D+l],s^A,\tau)-\int_{t'}^{t_F}d\tau\,\delta R_X^{ID}(s^D,s^A,\tau)\right\}\end{bmatrix},$$

--- Eq.(A1.2.1f)

and each type (g) history's contribution is:

$$\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D+l),\hat{M}_D(\Delta L^D+2,\Delta L^A+\Delta L^D+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$

$$=\int_{t_I}^{t_F}dt\int_t^{t_F}dt'\begin{bmatrix} r_I(1,\Delta L^D+l;s^A,t)\,r_D(\Delta L^D+2,\Delta L^A+\Delta L^D+l+1;s^A\cdot[1,+\Delta L^D+l],t') \\ \times\exp\left\{-\int_t^{t'}d\tau\,\delta R_X^{ID}(s^A\cdot[1,+\Delta L^D+l],s^A,\tau)-\int_{t'}^{t_F}d\tau\,\delta R_X^{ID}(s^D,s^A,\tau)\right\}\end{bmatrix}.$$

--- Eq.(A1.2.1g)

Under Dawg's indel model (or, similarly, under the "long indel" model), we have $\delta R_X^{ID}(s^D,s^A,\tau)=(\lambda_I+\lambda_D)(\Delta L^D-\Delta L^A)$, $\delta R_X^{ID}(s_0,s^A,\tau)=-(\lambda_I+\lambda_D)\Delta L^A$, and $\delta R_X^{ID}(s^A\cdot[\Delta L^A+1,+\Delta L^D+l],s^A,\tau)=\delta R_X^{ID}(s^A\cdot[1,+\Delta L^D+l],s^A,\tau)=(\lambda_I+\lambda_D)(\Delta L^D+l)$.
Using them, Eqs.(A1.2.1e,f,g) are calculated as:

$$\mu_P\left[\left(\left[\hat{M}_D(2,\Delta L^A+1),\hat{M}_I(1,\Delta L^D)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$

$$=\frac{\lambda_D f_D(\Delta L^A)\,\lambda_I f_I(\Delta L^D)}{(\lambda_I+\lambda_D)\Delta L^D}\left[\frac{e^{(\lambda_I+\lambda_D)\Delta L^A(t_F-t_I)}-1}{(\lambda_I+\lambda_D)\Delta L^A}-F_{e-}\left((\lambda_I+\lambda_D)(\Delta L^D-\Delta L^A),t_F-t_I\right)\right],$$

--- Eq.(A1.2.1e')

$$\mu_P\left[\left(\left[\hat{M}_I(\Delta L^A+1,\Delta L^D+l),\hat{M}_D(2,\Delta L^A+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$

$$=\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D+l),\hat{M}_D(\Delta L^D+2,\Delta L^A+\Delta L^D+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$

$$=\frac{\lambda_I f_I(\Delta L^D+l)\,\lambda_D f_D(\Delta L^A+l)}{(\lambda_I+\lambda_D)(\Delta L^A+l)}\left[F_{e-}\left((\lambda_I+\lambda_D)(\Delta L^D-\Delta L^A),t_F-t_I\right)-\frac{1-e^{-(\lambda_I+\lambda_D)(\Delta L^D+l)(t_F-t_I)}}{(\lambda_I+\lambda_D)(\Delta L^D+l)}\right].$$

--- Eqs.(A1.2.1f',g')

Here, we defined the function $F_{e-}(x,t)$ as:

$$F_{e-}(x,t)\equiv\begin{cases}\dfrac{1-e^{-xt}}{x} & if\quad x\neq 0,\\ t & if\quad x=0.\end{cases}\qquad \text{--- Eq.(A1.2.2)}$$

Meanwhile, each next-fewest-indel history is composed of three indel events, and classified into one of 6 broad types: (h) two successive deletions followed by an

insertion; (i) a deletion, followed by an insertion, followed by a deletion; (j) an insertion followed by two successive deletions; (k) a deletion followed by two successive insertions; (l) an insertion, followed by a deletion, followed by an insertion; and (m) two successive insertions followed by a deletion. And these 6 broad types can be further sub-classified into 24 sub-types, as explained in the following. Type (h) does not need to be sub-classified, and each history belonging to this type can be expressed as: $\left[\hat{M}_D(x, x+l-1), \hat{M}_D(2, \Delta L^A - l + 1), \hat{M}_I(1, \Delta L^D)\right]$ with $l = 1,..., \Delta L^A - 1$ and $x = 2,..., \Delta L^A - l + 2$. Type (i) is sub-classified into three sub-types: (i-1) the case where the insertion is immediately on the right of the sites to be deleted, represented as $\left[\hat{M}_D(x, x+l-1), \hat{M}_I(\Delta L^A - l + 1, \Delta L^D + l'), \hat{M}_D(2, \Delta L^A - l + l' + 1)\right]$ with $l = 1,..., \Delta L^A - 1$, $x = 2,..., \Delta L^A - l + 2$, and $l' = 0,...,\min\{L_I^{CO} - \Delta L^D, L_D^{CO} - \Delta L^A + l\}$; (i-2) the case where the insertion is immediately on the left of the sites to be deleted, represented as $\left[\hat{M}_D(x, x+l-1), \hat{M}_I(1, \Delta L^D + l'), \hat{M}_D(\Delta L^D + 2, \Delta L^A - l + \Delta L^D + l' + 1)\right]$ also with $l = 1,..., \Delta L^A - 1$, $x = 2,..., \Delta L^A - l + 2$, and $l' = 0,...,\min\{L_I^{CO} - \Delta L^D, L_D^{CO} - \Delta L^A + l\}$; and (i-3) the case where the first event deletes all unpreserved ancestral sites, represented as $\left[\hat{M}_D(2, \Delta L^A + 1), \hat{M}_I(1, \Delta L^D + l), \hat{M}_D(x, x+l-1)\right]$ with $l = 1,..., \min\{L_D^{CO}, L_I^{CO} - \Delta L^D\}$ and $x = 2,..., \Delta L^D + 2$. Type (j) is sub-classified into eight sub-types: (j-1) the case where the two deletions were on the right of the inserted sites that ended up in the descendant, represented as $\left[\hat{M}_I(\Delta L^A + 1, \Delta L^D + l), \hat{M}_D(x, x+l'-1), \hat{M}_D(2, \Delta L^A + l - l' + 1)\right]$ with $l = 0,..., L_I^{CO} - \Delta L^D$, $l' = \max\{1, \Delta L^A + l - L_D^{CO}\}, ..., \min\{L_D^{CO}, \Delta L^A + l - 1\}$, and $x = 2,..., \Delta L^A + l - l' + 2$; (j-2) the case where the two deletions were on the left of the inserted sites that ended up in the descendant, represented as $\left[\hat{M}_I(1, \Delta L^D + l), \hat{M}_D(x, x+l'-1), \hat{M}_D(\Delta L^D + 2, \Delta L^A + \Delta L^D + l - l' + 1)\right]$ with $l = 0,..., L_I^{CO} - \Delta L^D$, $l' = \max\{1, \Delta L^A + l - L_D^{CO}\}, ..., \min\{L_D^{CO}, \Delta L^A + l - 1\}$, and $x = \Delta L^D + 2,..., \Delta L^A + \Delta L^D + l - l' + 2$; (j-3) the case where the first and the second deletions were on the left and in the middle, respectively, of the inserted sites that ended up in the descendant, represented as $\left[\hat{M}_I(\Delta L^A + 1, \Delta L^D + l), \hat{M}_D(2, \Delta L^A + l' + 1), \hat{M}_D(x, x+l-l'-1)\right]$ with $l = 1,..., L_I^{CO} - \Delta L^D$, $l' = \max\{0, l - L_D^{CO}\}, ..., \min\{l - 1, L_D^{CO} - \Delta L^A\}$, and $x = 3,..., \Delta L^D + 1$ ($\Delta L^D \geq 2$ must always hold); (j-4) the case where the first and the second deletions were in the middle and on the left, respectively, of the inserted sites that ended up in the descendant, represented as $\left[\hat{M}_I(\Delta L^A + 1, \Delta L^D + l), \hat{M}_D(x, x+l-l'-1), \hat{M}_D(2, \Delta L^A + l' + 1)\right]$ with $l = 1,..., L_I^{CO} - \Delta L^D$, $l' = \max\{0, l - L_D^{CO}\}, ..., \min\{l - 1, L_D^{CO} - \Delta L^A\}$, and $x = \Delta L^A + l' + 3,..., \Delta L^A + \Delta L^D + l' + 1$ ($\Delta L^D \geq 2$ must always hold); (j-5) the case where the first and the second deletions were on the right and in the middle, respectively, of the inserted sites that ended up in the descendant, represented as $\left[\hat{M}_I(1, \Delta L^D + l), \hat{M}_D(\Delta L^D + l - l' + 2, \Delta L^A + \Delta L^D + l + 1), \hat{M}_D(x, x+l-l'-1)\right]$ with

$l = 1,..., L_I^{CO} - \Delta L^D$, $l' = \max\{0, l - L_D^{CO}\}, ..., \min\{l-1, L_D^{CO} - \Delta L^A\}$, and $x = 3,..., \Delta L^D + 1$ ($\Delta L^D \geq 2$ must always hold); (j-6) the case where the first and second deletions were in the middle and on the right, respectively, of the inserted sites that ended up in the descendant, represented as $\left[\hat{M}_I(1, \Delta L^D + l), \hat{M}_D(x, x + l - l' - 1), \hat{M}_D(\Delta L^D + 2, \Delta L^A + \Delta L^D + l' + 1)\right]$ with $l = 1,..., L_I^{CO} - \Delta L^D$, $l' = \max\{0, l - L_D^{CO}\}, ..., \min\{l-1, L_D^{CO} - \Delta L^A\}$, and $x = 3,..., \Delta L^D + 1$ ($\Delta L^D \geq 2$ must always hold); (j-7) the case the first and the second deletions were on the right and on the left, respectively, of the inserted sites that ended up in the descendant, represented as $\left[\hat{M}_I(x+1, \Delta L^D + l), \hat{M}_D(\Delta L^D + l + l'' + 2, \Delta L^A + \Delta L^D + l + 1), \hat{M}_D(2, l + l'' + 1)\right]$ with $l = 0, ..., L_I^{CO} - \Delta L^D$, $l'' = \max\{1 - l, \Delta L^A - L_D^{CO}\},..., \min\{\Delta L^A - 1, L_D^{CO} - l\}$, and $x = \max\{0, l''\},..., \min\{l + l'', \Delta L^A\}$; and (j-8) the case where the first and the second deletions were on the left and on the right, respectively, of the inserted sites that ended up in the descendant, represented as $\left[\hat{M}_I(x+1, \Delta L^D + l), \hat{M}_D(2, l + l'' + 1), \hat{M}_D(\Delta L^D + 2, \Delta L^A + \Delta L^D - l'' + 1)\right]$ with $l = 0, ..., L_I^{CO} - \Delta L^D$, $l'' = \max\{1 - l, \Delta L^A - L_D^{CO}\},..., \min\{\Delta L^A - 1, L_D^{CO} - l\}$, and $x = \max\{0, l''\},..., \min\{l + l'', \Delta L^A\}$. Type (k) does not need be sub-classified, and each history belonging to this type can be represented as: $\left[\hat{M}_D(2, \Delta L^A + 1), \hat{M}_I(1, l), \hat{M}_I(x, \Delta L^D - l)\right]$ with $l = 1,..., \Delta L^D - 1$ and $x = 1,..., l + 1$. Type (l) is sub-classified into three sub-types: (l-1) the case where the first insertion was on the immediate right of the left-flanking PAS, and at least a site inserted by the event survived, represented as $\left[\hat{M}_I(1, \Delta L^D + l - l'), \hat{M}_D(\Delta L^D - l' + 2, \Delta L^A + \Delta L^D + l - l' + 1), \hat{M}_I(x, l')\right]$ with $l = 0,..., L_D^{CO} - \Delta L^A$, $l' = \max\{1, \Delta L^D + l - L_I^{CO}\}, ..., \min\{\Delta L^D - 1, L_I^{CO}\}$, and $x = 1,..., \Delta L^D - l' + 1$; (l-2) the case where the first insertion was on the immediate left of the right-flanking PAS, and at least a site inserted by the event survived, represented as $\left[\hat{M}_I(\Delta L^A + 1, \Delta L^D + l - l'), \hat{M}_D(2, \Delta L^A + l + 1), \hat{M}_I(x, l')\right]$ with $l = 0,..., L_D^{CO} - \Delta L^A$, $l' = \max\{1, \Delta L^D + l - L_I^{CO}\}, ..., \min\{\Delta L^D - 1, L_I^{CO}\}$, and $x = 1,..., \Delta L^D - l' + 1$; and (l-3) the case where all sites inserted by the first event were deleted along with the unpreserved ancestral sites, represented as $\left[\hat{M}_I(x, l), \hat{M}_D(2, \Delta L^A + l + 1), \hat{M}_I(1, \Delta L^D)\right]$ with $l = 1,..., \min\{L_I^{CO}, L_D^{CO} - \Delta L^A\}$ and $x = 1,..., \Delta L^A + 1$. Finally, type (m) is sub-classified into eight sub-types: (m-1) the case where the two insertions are on the left of the unpreserved ancestral sites, represented as: $\left[\hat{M}_I(1, \Delta L^D + l - l'), \hat{M}_I(x, l'), \hat{M}_D(\Delta L^D + 2, \Delta L^A + \Delta L^D + l + 1)\right]$ with $l = 0,..., L_D^{CO} - \Delta L^A$, $l' = \max\{1, \Delta L^D + l - L_I^{CO}\}, ..., \min\{\Delta L^D + l - 1, L_I^{CO}\}$, and $x = 1,..., \Delta L^D + l - l' + 1$; (m-2) the case where the two insertions are on the right of the unpreserved ancestral sites, represented as $\left[\hat{M}_I(\Delta L^A + 1, \Delta L^D + l - l'), \hat{M}_I(x, l'), \hat{M}_D(2, \Delta L^A + l + 1)\right]$ with $l = 0,..., L_D^{CO} - \Delta L^A$, $l' = \max\{1, \Delta L^D + l - L_I^{CO}\}, ..., \min\{\Delta L^D + l - 1, L_I^{CO}\}$, and

$x = \Delta L^A + 1, ..., \Delta L^A + \Delta L^D + l - l' + 1$; (m-3) the case where the first and the second insertions are in the middle and on the left, respectively, of the unpreserved ancestral sites, represented as $\left[\hat{M}_I(x, l-l'), \hat{M}_I(1, \Delta L^D + l'), \hat{M}_D(\Delta L^D + 2, \Delta L^A + \Delta L^D + l + 1)\right]$ with $x = 2, ..., \Delta L^A$, $l = 1, ..., L_D^{CO} - \Delta L^A$, and $l' = \max\{0, l - L_I^{CO}\}, ..., \min\{l-1, L_I^{CO} - \Delta L^D\}$ ($\Delta L^A \geq 2$ must always hold); (m-4) the case where the first and the second insertions are on the left and in the middle, respectively, of the unpreserved ancestral sites, represented as $\left[\hat{M}_I(1, \Delta L^D + l'), \hat{M}_I(x, l-l'), \hat{M}_D(\Delta L^D + 2, \Delta L^A + \Delta L^D + l + 1)\right]$ with $l = 1, ..., L_D^{CO} - \Delta L^A$, $l' = \max\{0, l - L_I^{CO}\}, ..., \min\{l-1, L_I^{CO} - \Delta L^D\}$, and $x = \Delta L^D + l' + 2, ..., \Delta L^A + \Delta L^D + l'$ ($\Delta L^A \geq 2$ must always hold); (m-5) the case where the first and the second insertions are in the middle and on the right, respectively, of the unpreserved ancestral sites, represented as $\left[\hat{M}_I(x, l-l'), \hat{M}_I(\Delta L^A + l - l' + 1, \Delta L^D + l'), \hat{M}_D(2, \Delta L^A + l + 1)\right]$ with $x = 2, ..., \Delta L^A$, $l = 1, ..., L_D^{CO} - \Delta L^A$, and $l' = \max\{0, l - L_I^{CO}\}, ..., \min\{l-1, L_I^{CO} - \Delta L^D\}$ ($\Delta L^A \geq 2$ must always hold); (m-6) the case where the first and the second insertions are on the right and in the middle, respectively, of the unpreserved ancestral sites, represented as $\left[\hat{M}_I(\Delta L^A + 1, \Delta L^D + l'), \hat{M}_I(x, l-l'), \hat{M}_D(2, \Delta L^A + l + 1)\right]$ with $x = 2, ..., \Delta L^A$, $l = 1, ..., L_D^{CO} - \Delta L^A$, and $l' = \max\{0, l - L_I^{CO}\}, ..., \min\{l-1, L_I^{CO} - \Delta L^D\}$ ($\Delta L^A \geq 2$ must always hold); (m-7) the case where the first and the second insertions are on the left and on the right, respectively, of the unpreserved ancestral sites, represented as $\left[\hat{M}_I(1, l + l''), \hat{M}_I(\Delta L^A + l + l'' + 1, \Delta L^D - l''), \hat{M}_D(x + 2, \Delta L^A + l + x + 1)\right]$ with $l = 0, ..., L_D^{CO} - \Delta L^A$, $l'' = \max\{1 - l, \Delta L^D - L_I^{CO}\}, ..., \min\{\Delta L^D - 1, L_I^{CO} - l\}$, and $x = \max\{0, l''\}, ..., \min\{l + l'', \Delta L^D\}$; and (m-8) the case where the first and the second insertions are on the right and on the left, respectively, of the unpreserved ancestral sites, represented as $\left[\hat{M}_I(\Delta L^A + 1, \Delta L^D - l''), \hat{M}_I(1, l + l''), \hat{M}_D(x + 2, \Delta L^A + l + x + 1)\right]$ with $l = 0, ..., L_D^{CO} - \Delta L^A$, $l'' = \max\{1 - l, \Delta L^D - L_I^{CO}\}, ..., \min\{\Delta L^D - 1, L_I^{CO} - l\}$, and $x = \max\{0, l''\}, ..., \min\{l + l'', \Delta L^D\}$.

Using this classification of the next-fewest-indel histories in case (iv), the sum of their contributions can be expressed as:

$$\tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa = 3; \gamma_\kappa; \alpha(s^A, s^D)\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]$$
$$= \tilde{\mu}_P[(h)] + \sum_{\nu=1}^{3} \tilde{\mu}_P[(i-\nu)] + \sum_{\nu=1}^{8} \tilde{\mu}_P[(j-\nu)] + \tilde{\mu}_P[(k)] + \sum_{\nu=1}^{3} \tilde{\mu}_P[(l-\nu)] + \sum_{\nu=1}^{8} \tilde{\mu}_P[(m-\nu)] . \quad \text{--- Eq.(A1.2.3)}$$

Here, each of $\tilde{\mu}_P[(h)]$,…, $\tilde{\mu}_P[(m-\nu)]$ ($\nu = 1, ..., 8$) is the sum of contributions from the histories belonging to one of the 24 sub-types explained above. Each of them can be calculated by calculating the contribution from each constituent history according to the definition, Eq.(4.1.1b) of part I supplemented by Eq.(3.1.8b) of part I, and by summing the contributions over possible values of the length variables l and l' (or l'') and over possible positions x when necessary, all specified above. Under a space- and time-homogeneous model, the indel rates become independent of the positions of

the indels and of the exact state before the event, and the exit rate depends only on the sequence length (and possibly on the functional form of the deletion length distribution). This simplifies the calculation considerably. Especially, because each term is independent of the exact position of indels, a summation over the positions x is reduced to a simple multiplication by the number of possible positions. In addition, considerations on spatial symmetry reveal that some different sub-types actually give identical contributions to each other (detailed below). Moreover, under Dawg's indel model or the "long indel" model, the increment of the exit rate is proportional to the difference in the sequence lengths, which further simplifies the calculation. Here, we give the results of contributions by individual histories under Dawg's indel model. But the results apply immediately to, *e.g.*, the "long indel" model as well. We first give the general formula for the contribution of a three-event local history, and then we apply the formula to the histories of different sub-types.

First we consider a general three-event history, $\left[\hat{M}_1, \hat{M}_2, \hat{M}_3\right]$, on the ancestral state $s^A = \left[L, 1, ..., \Delta L^A, R\right]$, that resulted in the descendant state $s^D = \left[L, \upsilon_1^D, ..., \upsilon_{\Delta L^D}^D, R\right]$. Here $\hat{M}_\nu$ (with $\nu = 1,2,3$) is either an insertion ($\hat{M}_I(x,l)$) or a deletion ($\hat{M}_D(x_B, x_E)$). Let δl_ν (with $\nu = 1,2,3$) be the sequence length change caused by the event $\hat{M}_\nu$; δl_ν is positive if $\hat{M}_\nu$ is an insertion and negative if $\hat{M}_\nu$ is a deletion. They satisfy: $\delta l_1 + \delta l_2 + \delta l_3 = \Delta L_D - \Delta L_A$. And let $r(\delta l)$ be the space- and time-homogeneous rate of the indel event that changes the sequence length by δl. Under Dawg's indel model, we have:

$$r(\delta l) = \begin{cases} \lambda_I f_I(\delta l) & if \quad \delta l > 0, \\ \lambda_D f_D(-\delta l) & if \quad \delta l < 0. \end{cases} \quad \text{--- Eq.(A1.2.4)}$$

Then, according to Eq.(4.1.1b) of part I supplemented by Eq.(3.1.8b) of part I, the contribution from this three-event local history is expressed as:

$$\mu_P\left(\delta l_1, \delta l_2, \delta l_3; t_F - t_I\right) \equiv \mu_P\left[\left(\left[\hat{M}_1, \hat{M}_2, \hat{M}_3\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]\Bigg|_{Dawg} \text{ ---}$$

$$= r(\delta l_1) r(\delta l_2) r(\delta l_3) I_{3\hat{M}}\left(\delta l_1, \delta l_2, \delta l_3; t_F - t_I; \lambda_I + \lambda_D\right).$$

Eq.(A1.2.5a)

Here, $I_{3\hat{M}}\left(\delta l_1, \delta l_2, \delta l_3; t_F - t_I; \lambda_I + \lambda_D\right)$ is a triple-time integral:

$$I_{3\hat{M}}\left(\delta l_1, \delta l_2, \delta l_3; t_F - t_I; \lambda_I + \lambda_D\right)$$

$$\equiv \int_{t_I}^{t_F} dt_1 \int_{t_1}^{t_F} dt_2 \int_{t_2}^{t_F} dt_3 \exp\left\{\begin{array}{l} -(\lambda_I + \lambda_D)\delta l_1 (t_2 - t_1) - (\lambda_I + \lambda_D)(\delta l_1 + \delta l_2)(t_3 - t_2) \\ \quad -(\lambda_I + \lambda_D)(\delta l_1 + \delta l_2 + \delta l_3)(t_F - t_3) \end{array}\right\}$$

$$= \int_{t_I}^{t_F} dt_1 \int_{t_1}^{t_F} dt_2 \int_{t_2}^{t_F} dt_3\, e^{-(\lambda_I + \lambda_D)\delta l_1 (t_F - t_1)} e^{-(\lambda_I + \lambda_D)\delta l_2 (t_F - t_2)} e^{-(\lambda_I + \lambda_D)\delta l_3 (t_F - t_3)} .$$

--- Eq.(A1.2.5b)

When $\delta l_2 + \delta l_3 \neq 0$, it is calculated as:

$$I_{3\hat{M}}\left(\delta l_1, \delta l_2, \delta l_3; t_F - t_I; \lambda_I + \lambda_D\right)$$

$$= \frac{1}{(\lambda_I + \lambda_D)^2}\left[\begin{array}{l} \dfrac{1}{\delta l_2(\delta l_2 + \delta l_3)} \dfrac{1 - e^{-(\lambda_I + \lambda_D)\delta l_1 (t_F - t_I)}}{(\lambda_I + \lambda_D)\delta l_1} - \dfrac{1}{\delta l_3 \delta l_2} F_{e-}\left((\lambda_I + \lambda_D)(\delta l_1 + \delta l_2), t_F - t_I\right) \\ \quad + \dfrac{1}{\delta l_3(\delta l_2 + \delta l_3)} F_{e-}\left((\lambda_I + \lambda_D)(\delta l_1 + \delta l_2 + \delta l_3), t_F - t_I\right) \end{array}\right].$$

--- Eq.(A1.2.5b')

Here the function $F_{e-}(x,t)$ was already defined in Eq.(SR-8B.2). When $\delta l_2 + \delta l_3 = 0$, the triple-time integral is calculated as:

$$I_{3\hat{M}}\left(\delta l_1, \delta l_2, \delta l_3; t_F - t_I; \lambda_I + \lambda_D\right)$$
$$= \frac{1}{(\lambda_I + \lambda_D)^2}\left[\begin{array}{l} \dfrac{1}{\delta l_3 \delta l_2}\left\{\dfrac{1 - e^{-(\lambda_I+\lambda_D)\delta l_1 (t_F - t_I)}}{(\lambda_I + \lambda_D)\delta l_1} - F_{e-}\left((\lambda_I + \lambda_D)(\delta l_1 + \delta l_2), t_F - t_I\right)\right\} \\ \quad - \dfrac{1}{\delta l_3 \delta l_1}\dfrac{1 - \left(1 + (\lambda_I + \lambda_D)\delta l_1 (t_F - t_I)\right) e^{-(\lambda_I+\lambda_D)\delta l_1 (t_F - t_I)}}{(\lambda_I + \lambda_D)\delta l_1} \end{array}\right].$$

--- Eq.(A1.2.5b")

When $\hat{M}_2$ does not overlap $\hat{M}_3$, there should be a corresponding pair of events, $\hat{M}_2'$ and $\hat{M}_3'$, that satisfy $\hat{M}_2\hat{M}_3 \sim \hat{M}_3'\hat{M}_2'$ as one of the binary equivalence relations (Eqs.(2.3.3a-d) of part I). Under Dawg's model, the effects of $\hat{M}_2$ and $\hat{M}_3$ are *partially* factorable if we consider the joint contribution from the two local histories, $\left[\hat{M}_1, \hat{M}_2, \hat{M}_3\right]$ and $\left[\hat{M}_1, \hat{M}_3', \hat{M}_2'\right]$. We will express this joint contribution as:

$$\mu_P^{(S23)}\left(\delta l_1, \{\delta l_2, \delta l_3\}; t_F - t_I\right)$$
$$\equiv \mu_P\left[\left(\left[\hat{M}_1, \hat{M}_2, \hat{M}_3\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]\Bigg|_{Dawg} + \mu_P\left[\left(\left[\hat{M}_1, \hat{M}_3', \hat{M}_2'\right], [t_I, t_F]\right) \middle| (s^A, t_I)\right]\Bigg|_{Dawg}$$
$$= r(\delta l_1) r(\delta l_2) r(\delta l_3) I_{\hat{M}+2\hat{M}}\left(\delta l_1, \{\delta l_2, \delta l_3\}; t_F - t_I; \lambda_I + \lambda_D\right).$$

--- Eq.(A1.2.6a)

Here, $I_{\hat{M}+2\hat{M}}\left(\delta l_1, \{\delta l_2, \delta l_3\}; t_F - t_I; \lambda_I + \lambda_D\right)$ is a triple-time integral:

$$I_{\hat{M}+2\hat{M}}\left(\delta l_1, \{\delta l_2, \delta l_3\}; t_F - t_I; \lambda_I + \lambda_D\right)$$
$$\equiv I_{3\hat{M}}\left(\delta l_1, \delta l_2, \delta l_3; t_F - t_I; \lambda_I + \lambda_D\right) + I_{3\hat{M}}\left(\delta l_1, \delta l_3, \delta l_2; t_F - t_I; \lambda_I + \lambda_D\right) \quad \text{--- Eq.(A1.2.6b)}$$
$$= \int_{t_I}^{t_F} dt_1 \left[e^{-(\lambda_I+\lambda_D)\delta l_1 (t_F - t_1)} \int_{t_1}^{t_F} dt_2\, e^{-(\lambda_I+\lambda_D)\delta l_2 (t_F - t_2)} \int_{t_1}^{t_F} dt_3\, e^{-(\lambda_I+\lambda_D)\delta l_3 (t_F - t_3)}\right].$$

It is calculated as:

$$I_{\hat{M}+2\hat{M}}\left(\delta l_1, \{\delta l_2, \delta l_3\}; t_F - t_I; \lambda_I + \lambda_D\right)$$
$$= \frac{1}{(\lambda_I + \lambda_D)^2 \delta l_2 \delta l_3}\left[\begin{array}{l} \dfrac{1 - e^{-(\lambda_I+\lambda_D)\delta l_1 (t_F - t_I)}}{(\lambda_I + \lambda_D)\delta l_1} + F_{e-}\left((\lambda_I + \lambda_D)(\delta l_1 + \delta l_2 + \delta l_3), t_F - t_I\right) \\ - F_{e-}\left((\lambda_I + \lambda_D)(\delta l_1 + \delta l_2), t_F - t_I\right) - F_{e-}\left((\lambda_I + \lambda_D)(\delta l_1 + \delta l_3), t_F - t_I\right) \end{array}\right].$$

--- Eq.(A1.2.6b')

Similarly, when $\hat{M}_1$ does not overlap $\hat{M}_2$, there should be a corresponding pair of events, $\hat{M}_1'$ and $\hat{M}_2'$, that satisfy the binary equivalence, $\hat{M}_1\hat{M}_2 \sim \hat{M}_2'\hat{M}_1'$ (as one of Eqs.(2.3.3a-d) of part I). When considering the joint contribution from the two local histories, $\left[\hat{M}_1, \hat{M}_2, \hat{M}_3\right]$ and $\left[\hat{M}_2', \hat{M}_1', \hat{M}_3\right]$, the effects of $\hat{M}_1$ and $\hat{M}_2$ are *partially* factorable under Dawg's model. We will express the joint effect as:

$$\mu_P^{(S12)}\left(\{\delta l_1,\delta l_2\},\delta l_3;t_F-t_I\right)$$
$$\equiv \mu_P\left[\left(\left[\ \hat{M}_1,\hat{M}_2,\hat{M}_3\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right]\Bigg|_{Dawg}+\mu_P\left[\left(\left[\ \hat{M}'_2,\hat{M}'_1,\hat{M}_3\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right]\Bigg|_{Dawg}$$
$$=r(\delta l_1)r(\delta l_2)r(\delta l_3)I_{2\hat{M}+\hat{M}}\left(\{\delta l_1,\delta l_2\},\delta l_3;t_F-t_I;\lambda_I+\lambda_D\right).$$

--- Eq.(A1.2.7a)

Here, $I_{2\hat{M}+\hat{M}}\left(\{\delta l_1,\delta l_2\},\delta l_3;t_F-t_I;\lambda_I+\lambda_D\right)$ is a triple-time integral:

$$I_{2\hat{M}+\hat{M}}\left(\{\delta l_1,\delta l_2\},\delta l_3;t_F-t_I;\lambda_I+\lambda_D\right)$$
$$\equiv I_{3\hat{M}}\left(\delta l_1,\delta l_2,\delta l_3;t_F-t_I;\lambda_I+\lambda_D\right)+I_{3\hat{M}}\left(\delta l_2,\delta l_1,\delta l_3;t_F-t_I;\lambda_I+\lambda_D\right)$$ --- Eq.(A1.2.7b)
$$=\int_{t_I}^{t_F}dt_3\left[e^{-(\lambda_I+\lambda_D)\delta l_3(t_F-t_3)}\int_{t_I}^{t_3}dt_1\,e^{-(\lambda_I+\lambda_D)\delta l_1(t_F-t_1)}\int_{t_I}^{t_3}dt_2\,e^{-(\lambda_I+\lambda_D)\delta l_2(t_F-t_2)}\right].$$

It could be calculated by taking advantage of the following relationship:

$$I_{2\hat{M}+\hat{M}}\left(\{\delta l_1,\delta l_2\},\delta l_3;t_F-t_I;\lambda_I+\lambda_D\right)$$
$$=e^{-(\lambda_I+\lambda_D)(\delta l_1+\delta l_2+\delta l_3)(t_F-t_I)}I_{\hat{M}+2\hat{M}}\left(-\delta l_3,\{-\delta l_2,-\delta l_1\};t_F-t_I;\lambda_I+\lambda_D\right).$$ --- Eq.(A1.2.7b')

Now we can apply the general formulas, Eqs(A1.2.5a,b"), Eqs.(A1.2.6a,b'), and Eqs.(A1.2.7a,b'), to specific cases. First, the contribution from a type (h) history is:

$$\mu_P\left[\left(\left[\ \hat{M}_D(x,x+l-1),\hat{M}_D(2,\Delta L^A-l+1),\hat{M}_I(1,\Delta L^D)\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right]$$ ---
$$=\mu_P\left(-l,-(\Delta L^A-l),+\Delta L^D;t_F-t_I\right).$$

Eq.(A1.2.8a)

Second, by spatial symmetry, the contribution from a type (i-1) history is identical to that from the corresponding type (i-2) history. They are given by:

$$\mu_P\left[\left(\left[\hat{M}_D(x,x+l-1),\hat{M}_I(\Delta L^A-l+1,\Delta L^D+l'),\hat{M}_D(2,\Delta L^A-l+l'+1)\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right]$$
$$=\mu_P\left[\left(\left[\hat{M}_D(x,x+l-1),\hat{M}_I(1,\Delta L^D+l'),\hat{M}_D(\Delta L^D+2,\Delta L^A-l+\Delta L^D+l'+1)\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right]$$
$$=\mu_P\left(-l,+(\Delta L^D+l'),-(\Delta L^A-l+l');t_F-t_I\right).$$

--- Eqs.(A1.2.8b,c)

Third, the contribution from a type (i-3) history is:

$$\mu_P\left[\left(\left[\hat{M}_D(2,\Delta L^A+1),\hat{M}_I(1,\Delta L^D+l),\hat{M}_D(x,x+l-1)\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right]$$ ---
$$=\mu_P\left(-\Delta L^A,+(\Delta L^D+l),-l;t_F-t_I\right).$$

Eq.(A1.2.8d)

Fourth, by spatial symmetry, the contribution from a type (j-1) history is identical to that from the corresponding type (j-2) history. They are given by:

$$\mu_P\left[\left(\left[\hat{M}_I(\Delta L^A+1,\Delta L^D+l),\hat{M}_D(x,x+l'-1),\hat{M}_D(2,\Delta L^A+l-l'+1)\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right]$$
$$=\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D+l),\hat{M}_D(x,x+l'-1),\hat{M}_D(\Delta L^D+2,\Delta L^A+\Delta L^D+l-l'+1)\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right]$$
$$=\mu_P\left(+(\Delta L^D+l),-l',-(\Delta L^A+l-l');t_F-t_I\right).$$

--- Eqs.(A1.2.8e,f)

Fifth, we consider the joint contribution from a type (j-3) history and the corresponding type (j-4) history. By spatial symmetry, it is identical to the joint contribution from the corresponding type (j-5) and type (j-6) histories. They are given by:

$$\mu_P\left[\left(\left[\hat{M}_I(\Delta L^A+1,\Delta L^D+l),\hat{M}_D(2,\Delta L^A+l'+1),\hat{M}_D(x,x+l-l'-1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$+\mu_P\left[\left(\left[\hat{M}_I(\Delta L^A+1,\Delta L^D+l),\hat{M}_D(x,x+l-l'-1),\hat{M}_D(2,\Delta L^A+l'+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D+l),\hat{M}_D(\Delta L^D+l-l'+2,\Delta L^A+\Delta L^D+l+1),\hat{M}_D(x,x+l-l'-1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$+\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D+l),\hat{M}_D(x,x+l-l'-1),\hat{M}_D(\Delta L^D+2,\Delta L^A+\Delta L^D+l'+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P^{(S23)}\left(+(\Delta L^D+l),\{-(\Delta L^A+l'),-(l-l')\};t_F-t_I\right).$$

--- Eqs.(A1.2.8g,h)

Sixth, consider the joint contribution from a type (j-7) history and the corresponding type (j-8) history. It is given by:

$$\mu_P\left[\left(\left[\hat{M}_I(x+1,\Delta L^D+l),\hat{M}_D(\Delta L^D+l+l''+2,\Delta L^A+\Delta L^D+l+1),\hat{M}_D(2,l+l''+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$+\mu_P\left[\left(\left[\hat{M}_I(x+1,\Delta L^D+l),\hat{M}_D(2,l+l''+1),\hat{M}_D(\Delta L^D+2,\Delta L^A+\Delta L^D-l''+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P^{(S23)}\left(+(\Delta L^D+l),\{-(\Delta L^A-l''),-(l+l'')\};t_F-t_I\right).$$

--- Eq.(A1.2.8i)

Seventh, the contribution from a type (k) history is:

$$\mu_P\left[\left(\left[\hat{M}_D(2,\Delta L^A+1),\hat{M}_I(1,l),\hat{M}_I(x,\Delta L^D-l)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P\left(-\Delta L^A,+l,+(\Delta L^D-l);t_F-t_I\right).$$

--- Eq.(A1.2.8j)

Eighth, by spatial symmetry, the contribution from a type (l-1) history is identical to that from the corresponding type (l-2) history. They are given by:

$$\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D+l-l'),\hat{M}_D(\Delta L^D-l'+2,\Delta L^A+\Delta L^D+l-l'+1),\hat{M}_I(x,l')\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P\left[\left(\left[\hat{M}_I(\Delta L^A+1,\Delta L^D+l-l'),\hat{M}_D(2,\Delta L^A+l+1),\hat{M}_I(x,l')\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P\left(+(\Delta L^D+l-l'),-(\Delta L^A+l),+l';t_F-t_I\right).$$

--- Eqs.(A1.2.8k,l)

Ninth, the contribution from a type (l-3) history is:

$$\mu_P\left[\left(\left[\hat{M}_I(x,l),\hat{M}_D(2,\Delta L^A+l+1),\hat{M}_I(1,\Delta L^D)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P\left(+l,-(\Delta L^A+l),+\Delta L^D;t_F-t_I\right).$$

--- Eq.(A1.2.8m)

Tenth, by spatial symmetry, the contribution from a type (m-1) history is identical to that from the corresponding type (m-2) history. They are given by:

$$\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D+l-l'),\hat{M}_I(x,l'),\hat{M}_D(\Delta L^D+2,\Delta L^A+\Delta L^D+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P\left[\left(\left[\hat{M}_I(\Delta L^A+1,\Delta L^D+l-l'),\hat{M}_I(x,l'),\hat{M}_D(2,\Delta L^A+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P\left(+(\Delta L^D+l-l'),+l',-(\Delta L^A+l);t_F-t_I\right).$$

--- Eqs.(A1.2.8n,o)

Eleventh, consider the joint contribution from a type (m-3) history and the corresponding type (m-4) history. By spatial symmetry, it is equal to the joint contribution from the corresponding type (m-5) and type (m-6) histories. They are given by:

$$\mu_P\left[\left(\left[\hat{M}_I(x,l-l'),\hat{M}_I(1,\Delta L^D+l'),\hat{M}_D(\Delta L^D+2,\Delta L^A+\Delta L^D+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$+\mu_P\left[\left(\left[\hat{M}_I(1,\Delta L^D+l'),\hat{M}_I(x,l-l'),\hat{M}_D(\Delta L^D+2,\Delta L^A+\Delta L^D+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P\left[\left(\left[\hat{M}_I(x,l-l'),\hat{M}_I(\Delta L^A+l-l'+1,\Delta L^D+l'),\hat{M}_D(2,\Delta L^A+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$+\mu_P\left[\left(\left[\hat{M}_I(\Delta L^A+1,\Delta L^D+l'),\hat{M}_I(x,l-l'),\hat{M}_D(2,\Delta L^A+l+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P^{(S12)}\left(\{+(l-l'),+(\Delta L^D+l')\},-(\Delta L^A+l);t_F-t_I\right).$$

--- Eqs.(A1.2.8p,q)

And, finally, twelfth, consider the joint contribution from a type (m-7) history and the corresponding type (m-8) history. It is given by:

$$\mu_P\left[\left(\left[\hat{M}_I(1,l+l''),\hat{M}_I(\Delta L^A+l+l''+1,\Delta L^D-l''),\hat{M}_D(x+2,\Delta L^A+l+x+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$+\mu_P\left[\left(\left[\hat{M}_I(\Delta L^A+1,\Delta L^D-l''),\hat{M}_I(1,l+l''),\hat{M}_D(x+2,\Delta L^A+l+x+1)\right],[t_I,t_F]\right)\middle|(s^A,t_I)\right]$$
$$=\mu_P^{(S12)}\left(\{+(l+l''),+(\Delta L^D-l'')\},-(\Delta L^A+l);t_F-t_I\right).$$

--- Eq.(A1.2.8r)

These equations, Eqs.(A1.2.8a-r), supplemented with the general formulas, Eqs.(A1.2.5,6,7), provide the elementary contributions from the individual next-fewest-indel local histories. By summing them up over the appropriate lengths and positions given above the general formulas, Eqs.(A1.2.5,6,7), we get the total contribution of the next-fewest-indel local histories.

A1.3. System of integral equations providing "exact" multiplication factors for case-(i) & (iii) local PWAs

In Section 1.2 of Results, we derived Eq.(1.2.7), which, supplemented with Eq.(1.2.8), gives a system of integral equations that can in principle provide "exact" multiplication factors for cases (i) and (ii), where there are a non-negative integer (ΔL^A) of ancestral sites but zero descendant sites in between a pair of PASs. Here, following a similar line of procedures, we will derive a system of integral equations that can in principle provide "exact" multiplication factors *for cases (i) and (iii)*, where there are zero ancestral sites but a non-negative integer (ΔL^D) of descendant sites in between a pair of PASs. For this purpose, we assume a setting very similar to that made above Eq.(1.2.6) for cases (i) and (ii). The only notable difference is that the ancestral and the descendant states here are $s^A=[L,R]$ and

$s^D = \left[L, \upsilon_1^D, ..., \upsilon_{\Delta L^D}^D, R\right]$, respectively, which will be denoted as $\left\langle s^A \right| = \left\langle 0 \right|$ and $\left\langle s^D \right| = \left\langle \Delta L^D \right|$ under the setting of (local) spatial homogeneity.

The starting point here is the fundamental integral equation, Eq.(3.1.2) of part I (Ezawa, Graur and Landan 2015a), for the stochastic evolution operator $\hat{P}^{ID}(t_I, t_F)$, instead of Eq.(3.1.4) of part I for cases (i) and (ii). Similarly to above Eq.(1.2.6), we sandwich Eq.(3.1.2) of part I with $\left\langle s^A \right|$ and $\left| s^D \right\rangle$, expand $\hat{Q}_M^{ID}(t)$ using its definition (*i.e.*, Eq.(3.1.1c) of part I supplemented with Eqs.(2.4.2b',c') of part I), and ignore the effects of indels that are irrelevant to the cases under consideration. Then, we get:

$$\left\langle s^A \middle| \hat{P}^{ID}(t_I, t_F) \middle| s^D \right\rangle = \left\langle s^A \middle| s^D \right\rangle \exp\left\{-\int_{t_I}^{t_F} dt\, R_X^{ID}\left(\Delta L^A, t\right)\right\}$$
$$+ \sum_{l=1}^{\min\{\Delta L^D, L_I^{CO}\}} \sum_{x=1}^{\Delta L^D - l + 1} \int_{t_I}^{t_F} dt \left[\left\langle s^A \middle| \hat{P}^{ID}(t_I, t) \hat{M}_I(x,l) \middle| s^D \right\rangle g_I(l,t) \exp\left\{-\int_t^{t_F} d\tau\, R_X^{ID}\left(\Delta L^D, \tau\right)\right\}\right]$$
$$+ \sum_{l=1}^{L_D^{CO}} \sum_{x=2}^{\Delta L^D + 2} \int_{t_I}^{t_F} dt \left[\left\langle s^A \middle| \hat{P}^{ID}(t_I, t) \hat{M}_D(x, x+l-1) \middle| s^D \right\rangle g_D(l,t) \exp\left\{-\int_t^{t_F} d\tau\, R_X^{ID}\left(\Delta L^D, \tau\right)\right\}\right].$$

--- Eq.(A1.3.1)

Here, when considering the domains of the summations, we used the fact that the ket-vectors $\hat{M}_I(x,l)\left| s^D \right\rangle$ and $\hat{M}_D(x, x+l-1)\left| s^D \right\rangle$ correspond uniquely to the states $\left\langle \Delta L^D - l \right|$ and $\left\langle \Delta L^D + l \right|$, respectively. Next, we take advantage of this fact again, take account of the local homogeneity, and use the notation, $P^{ID}\left(\Delta L \mapsto \Delta L'; [t, t']\right) \equiv \left\langle \Delta L \middle| \hat{P}^{ID}(t, t') \middle| \Delta L' \right\rangle$, introduced above Eq.(1.2.7). Then, we can rewrite Eq.(A1.3.1) as:

$$P^{ID}\left(0 \mapsto \Delta L^D; [t_I, t_F]\right) = \delta(0, \Delta L^D) \exp\left\{-\int_{t_I}^{t_F} dt\, R_X^{ID}\left(\Delta L^D = 0, t\right)\right\}$$
$$+ \sum_{l=1}^{\min\{\Delta L^D, L_I^{CO}\}} (\Delta L^D - l + 1) \int_{t_I}^{t_F} dt \left[P^{ID}\left(0 \mapsto \Delta L^D - l; [t_I, t]\right) g_I(l,t) \exp\left\{-\int_t^{t_F} d\tau\, R_X^{ID}\left(\Delta L^D, \tau\right)\right\}\right]$$
$$+ (\Delta L^D + 1) \sum_{l=1}^{L_D^{CO}} \int_{t_I}^{t_F} dt \left[P^{ID}\left(0 \mapsto \Delta L^D + l; [t_I, t]\right) g_D(l,t) \exp\left\{-\int_t^{t_F} d\tau\, R_X^{ID}\left(\Delta L^D, \tau\right)\right\}\right].$$

--- Eq.(A1.3.2)

This gives the system of integral equations for the "exact" probabilities, $P^{ID}\left(0 \mapsto \Delta L^D; [t_I, t_F]\right)$, with non-negative integers $\Delta L^D = 0, 1, 2, ...$. This system of equations can be solved iteratively, again starting with the "zero-event approximation," $P_{\langle 0 \rangle}^{ID}\left(0 \mapsto \Delta L^D; [t_I, t_F]\right) = \delta(0, \Delta L^D) \exp\left\{-\int_{t_I}^{t_F} dt\, R_X^{ID}\left(\Delta L^D = 0, t\right)\right\}$. After a desired number (say, N_{ID}) of iteration steps, the multiplication factor will be obtained similarly, and we have:

$$
\begin{aligned}
&\tilde{\mu}_P^{\langle N_{ID}\rangle}\left[\left(\tilde{\Lambda}^{ID}\left[\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right] \\
&\equiv \sum_{N_\kappa=0}^{N_{ID}} \tilde{\mu}_P\left[\left(\Lambda^{ID}\left[N_\kappa;\gamma_\kappa;\alpha(s^A,s^D)\right],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right] \\
&= P_{\langle N_{ID}\rangle}^{ID}\left(0\mapsto \Delta L^D;[t_I,t_F]\right)\Big/ P\left[\left([],[t_I,t_F]\right)\middle|\,(s^A,t_I)\right] \\
&= \exp\left\{+\int_{t_I}^{t_F} dt\, R_X^{ID}\left(\Delta L^A=0,t\right)\right\} P_{\langle N_{ID}\rangle}^{ID}\left(0\mapsto \Delta L^D;[t_I,t_F]\right).
\end{aligned}
\qquad \text{--- Eq.(A1.3.3)}
$$

It should be noted that the exponent on the right-most hand side is the time integral of the exit rate of the ancestral state but not that of the descendant state. This is just due to the definition of the multiplication factor (Eq.(4.1.1b) of part I). The consideration of the time- and space-complexities of a naïve implementation of the iteration algorithm goes just as below Eq.(1.2.8) of Results. Just as Eq.(1.2.7), Eq.(A1.3.2) could also be numerically solved via a system of "two-sub-step" recursion relations:

$$
\begin{aligned}
P_{\langle n_S\rangle}^{ID}\left(0\mapsto \Delta L^D;[t_I,t]\right) = {}& \delta(0,\Delta L^D)\exp\left\{-\int_{t_I}^{t} d\tau\, R_X^{ID}\left(\Delta L^D=0,\tau\right)\right\} \\
&+\int_{t_I}^{t} dt'\left[\Phi_{\langle n_S\rangle}^{ID}\left(0\mapsto \Delta L^D;[t_I,t']\right)\exp\left\{-\int_{t'}^{t} d\tau\, R_X^{ID}\left(\Delta L^D,\tau\right)\right\}\right],
\end{aligned}
\qquad \text{--- Eq.(A1.3.4a)}
$$

and

$$
\begin{aligned}
\Phi_{\langle n_S\rangle}^{ID}\left(0\mapsto \Delta L^D;[t_I,t]\right) \equiv {}& \sum_{l=1}^{\min\{\Delta L^D,\, L_I^{CO}\}}\left[(\Delta L^D-l+1)P_{\langle n_S-1\rangle}^{ID}\left(0\mapsto \Delta L^D-l;[t_I,t]\right)g_I(l,t)\right] \\
&+(\Delta L^D+1)\sum_{l=1}^{L_D^{CO}}\left[P_{\langle n_S-1\rangle}^{ID}\left(0\mapsto \Delta L^D+l;[t_I,t]\right)g_D(l,t)\right].
\end{aligned}
\qquad \text{--- Eq.(A1.3.4b)}
$$

Similarly to Eqs.(1.2.9a,b), the algorithm to numerically solve this system of recursion relations also has the time-complexity of $O\left(N_{ID}\,L^{CO}(L^{CO}+N_P)N_P\right)$ and the space-complexity of $O\left(L^{CO}N_P\right)$, and a single run of the algorithm (with N_{ID} steps) outputs $P_{\langle N_{ID}\rangle}^{ID}\left(0\mapsto \Delta L^D;[t_I,t]\right)$ for all $\Delta L^D=0,1,2,\ldots,\Delta L^D_{\max}(\le L^{CO})$ and at all time-points, $t=t_I+i\,\frac{t_F-t_I}{N_P}$ with $i=0,1,\ldots,N_P$.

A2. Perturbation calculation of multiplication factors for MSAs: case (IV) in Section 1.3

Here we detail the calculation of the portion of the multiplication factor contributed by the fewest-indel events as well as that contributed by the next-fewest-indel events in case (IV). This case was considered in Section 1.3 of Results, and its situation is illustrated in Figure 4 E. In this case, the external sequence states were represented as: $s_1=\left[L,1,\ldots,\Delta L^{D1},R\right]$, $s_2=\left[L,1,\ldots,i,j,\ldots,\Delta L^{D1},R\right]$, and $s_3=\left[L,R\right]$, with $1\le i+1<j\le \Delta L^{D1}+1$ but $(i,j)\ne(0,\Delta L^{D1}+1)$. Also in this case, $N_{\min}\left[C_{\mathrm{K}};\alpha[s_1,s_2,s_3];T\right]=2$, and there were two fewest-indel local histories. One, which we call type (E) here, starts with the root state $s^{Root}=s_0^{Root}\ (=s_2)$. It is represented by Eq.(1.3.14a), *i.e.*,

$$\left\{\vec{\hat{M}}(b_1)=\left[\hat{M}_I(i+1,\, j-i-1)\right],\, \vec{\hat{M}}(b_2)=\left[\ \ \right],\, \vec{\hat{M}}(b_3)=\left[\hat{M}_D(2,\, \Delta L^{D1}-j+i+2)\right]\right\}.$$

--- Eq.(A2.1a)

And the other, which we call type (F) here, starts with the root state $s^{Root}=s_1=\left[L,1,...,\Delta L^{D1},R\right]$, instead of s_0^{Root} $(=s_2)$. It is represented by Eq.(1.3.14b), *i.e.*,

$$\left\{\vec{\hat{M}}(b_1)=\left[\ \ \right],\, \vec{\hat{M}}(b_2)=\left[\hat{M}_D(i+2,\, j)\right],\, \vec{\hat{M}}(b_3)=\left[\hat{M}_D(2,\, \Delta L^{D1}+1)\right]\right\}.$$ --- Eq.(A2.1b)

Thus the portion of the multiplication factor due to the fewest-indel histories can be written as:

$$\tilde{\tilde{\mathrm{M}}}_P\left[\Lambda_\Psi^{ID}\left[N_\mathrm{K}=2;C_\mathrm{K};\alpha[s_1,s_2,s_3];T\right]\middle|\, T\right]=\tilde{\tilde{\mathrm{M}}}_P[(E)]+\tilde{\tilde{\mathrm{M}}}_P[(F)].$$ --- Eq.(A2.2a)

Here $\tilde{\tilde{\mathrm{M}}}_P[(E)]$ and $\tilde{\tilde{\mathrm{M}}}_P[(F)]$ are the contributions from the local histories of type (E) and type (F), respectively. These contributions can be calculated according to the definition, Eq.(1.1.2b) supplemented by Eqs.(4.2.4b,6b,8) of part I (Ezawa, Graur and Landan 2015a), similarly to the calculation of $\tilde{\tilde{\mathrm{M}}}_P[(3rd)]$ in case (III). First, the contribution from the type (E) history is:

$$\tilde{\tilde{\mathrm{M}}}_P[(E)]=\mu_P\left[s^{Root}=s_2,\, s_0^{Root}=s_2,\, n^{Root};C_\mathrm{K}\right]\exp\left\{-\sum_{m=1,2,3}\int_{t_I}^{t_{F:m}}dt\;\delta R_{X:m}^{ID}(s^{Root}=s_2,\, s_0^{Root}=s_2,t)\Big|_{b_m}\right\}$$

$$\times\int_{t_I}^{t_{F:1}}dt\left[r_{I:1}(i+1,\, j-i-1;\, s^{Root}=s_2,t)\exp\left\{-\int_t^{t_{F:1}}d\tau\;\delta R_{X:1}^{ID}(s_1,\, s^{Root}=s_2,t)\Big|_{b_1}\right\}\right]$$

$$\times\int_{t_I}^{t_{F:3}}dt\left[r_{D:3}(2,\Delta L^{D1}-j+i+2;\, s^{Root}=s_2,t)\exp\left\{-\int_t^{t_{F:3}}d\tau\;\delta R_{X:3}^{ID}(s_3,\, s^{Root}=s_2,t)\Big|_{b_3}\right\}\right].$$

--- Eq.(A2.2b)

And the contribution from the type (F) history is:

$$\tilde{\tilde{\mathrm{M}}}_P[(F)]=\mu_P\left[s^{Root}=s_1,\, s_0^{Root}=s_2,\, n^{Root};C_\mathrm{K}\right]\exp\left\{-\sum_{m=1,2,3}\int_{t_I}^{t_{F:m}}dt\;\delta R_{X:m}^{ID}(s^{Root}=s_1,\, s_0^{Root}=s_2,t)\Big|_{b_m}\right\}$$

$$\times\int_{t_I}^{t_{F:2}}dt\left[r_{D:2}(i+2,\, j;\, s^{Root}=s_1,t)\exp\left\{-\int_t^{t_{F:2}}d\tau\;\delta R_{X:2}^{ID}(s_2,\, s^{Root}=s_1,t)\Big|_{b_2}\right\}\right]$$

$$\times\int_{t_I}^{t_{F:3}}dt\left[r_{D:3}(2,\Delta L^{D1}+1;\, s^{Root}=s_2,t)\exp\left\{-\int_t^{t_{F:3}}d\tau\;\delta R_{X:3}^{ID}(s_3,\, s^{Root}=s_1,t)\Big|_{b_3}\right\}\right].$$

--- Eq.(A2.2c)

Assuming a uniform distribution of the ancestral sequence length, we have $\mu_P\left[s^{Root},\, s_0^{Root},\, n^{Root};C_\mathrm{K}\right]=1$ no matter what the root state (s^{Root}) is. Then, under Dawg's model, Eqs.(SR-9.2b,c) can be calculated as:

$$\tilde{\mathrm{M}}_P[(E)] = \lambda_{I:1} f_{I:1}(j-i-1)\frac{1-\exp\left(-(\lambda_{I:1}+\lambda_{D:1})(j-i-1)(t_{F:1}-t_I)\right)}{(\lambda_{I:1}+\lambda_{D:1})(j-i-1)}$$

$$\times \lambda_{D:3} f_{D:3}(\Delta L^{D1}-j+i+1)\frac{\exp\left(+(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)(t_{F:3}-t_I)\right)-1}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)},$$

--- Eq.(A2.2b')

and

$$\tilde{\mathrm{M}}_P[(F)] = \exp\left(-(\lambda_{I:1}+\lambda_{D:1})(j-i-1)(t_{F:1}-t_I)\right)$$

$$\times \lambda_{D:2} f_{D:2}(j-i-1)\frac{1-\exp\left(-(\lambda_{I:2}+\lambda_{D:2})(j-i-1)(t_{F:2}-t_I)\right)}{(\lambda_{I:2}+\lambda_{D:2})(j-i-1)}$$

$$\times \lambda_{D:3} f_{D:3}(\Delta L^{D1})\frac{\exp\left(+(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)(t_{F:3}-t_I)\right)-\exp\left(-(\lambda_{I:3}+\lambda_{D:3})(j-i-1)(t_{F:3}-t_I)\right)}{(\lambda_{I:3}+\lambda_{D:3})\Delta L^{D1}}.$$

--- Eq.(A2.2c')

The next-fewest-indel local histories consist of three indels each, and can be broadly classified into four types according to the root sequence state: (G) the histories with $s^{Root} = s_0^{Root}$ $(= s_2 = \left[L, 1, ..., i, j, ..., \Delta L^{D1}, R\right])$; (H) those with $s^{Root} = s_1 = \left[L, 1, ..., \Delta L^{D1}, R\right]$; (I) the histories having the root states made from s_1 by removing a single connected fraction out of the run of sites missing in s_2, *i.e.*, $s^{Root} = \left[L, 1, ..., i', j', ..., \Delta L^{D1}, R\right]$(with $i' \geq i$, $i'+2 \leq j' \leq j$, but $(i', j') \neq (i, j)$); and (J) the histories having the root states made by adding some extra sites to s_1, *i.e.*, $s^{Root} = \left[L, 1, ..., k, \upsilon'_1, ..., \upsilon'_l, k+1, ..., \Delta L^{D1}, R\right]$ with $k = i, ..., j-1$, $l = 1, ..., \min\{L^{CO}_{D:1}, L^{CO}_{D:2}-(j-i-1), L^{CO}_{D:3}-\Delta L^{D1}\}$, $\upsilon'_a \notin \{L, 1, ..., \Delta L^{D1}, R\}$ for $^\forall a \in \{1, ..., l\}$, and $\upsilon'_a \neq \upsilon'_b$ for $^\forall a \neq b\ (\in \{1, ..., l\})$. These four broad types are further sub-classified into 10 sub-types in total.

First, type (G) local histories are sub-classified into four sub-types: (G-1) the case where two successive insertions occur along b_1, represented as

$$\left\{\vec{\hat{M}}(b_1) = \left[\hat{M}_I(i+1, l), \hat{M}_I(x, j-i-1-l)\right], \vec{\hat{M}}(b_2) = [\ \ ], \vec{\hat{M}}(b_3) = \left[\hat{M}_D(2, \Delta L^{D1}-j+i+2)\right]\right\}$$

with $l = 1, ..., j-i-2$ and $x = i+1, ..., i+l+1$ ($j-i \geq 3$ must hold); (G-2) the case where a
"long" insertion and a subsequent deletion occur along b_1, represented as

$$\left\{\vec{\hat{M}}(b_1) = \left[\hat{M}_I(i+1, j-i-1+l), \hat{M}_D(x, x+l-1)\right], \vec{\hat{M}}(b_2) = [\ \ ], \vec{\hat{M}}(b_3) = \left[\hat{M}_D(2, \Delta L^{D1}-j+i+2)\right]\right\}$$

with $l = 1, ..., \min\{L^{CO}_{D:1}, L^{CO}_{I:1}-j+i+1\}$ and $x = i+2, ..., j+1$; (G-3) the case where two successive deletions occur along b_3, represented as

$$\left\{\vec{\hat{M}}(b_1) = \left[\hat{M}_I(i+1, j-i-1)\right], \vec{\hat{M}}(b_2) = [\ \ ], \vec{\hat{M}}(b_3) = \left[\hat{M}_D(x, x+l-1), \hat{M}_D(2, \Delta L^{D1}-j+i-l+2)\right]\right\}$$

with $l = 1,...,\Delta L^{D1} - j + i$ and $x = 2,...,\Delta L^{D1} - j + i - l + 3$ ($\Delta L^{D1} - j + i \geq 1$ must hold); and (G-4) the case where an insertion and a subsequent "long" deletion occur along b_3, represented as

$$\left\{\vec{\hat{M}}(b_1) = \left[\hat{M}_I(i+1, j-i-1)\right], \vec{\hat{M}}(b_2) = [\ ], \vec{\hat{M}}(b_3) = \left[\hat{M}_I(x,l), \hat{M}_D(2, \Delta L^{D1} - j + i + l + 2)\right]\right\}$$

with $l = 1,...,\min\{L^{CO}_{I:3}, L^{CO}_{D:3} - (\Delta L^{D1} - j + i + 1)\}$ and $x = 1,...,\Delta L^{D1} - j + i + 2$.

Second, type (H) local histories are also sub-classified into four sub-types: (H-1) the case where two successive deletions occur along b_2, represented as

$$\left\{\vec{\hat{M}}(b_1) = [\ ], \vec{\hat{M}}(b_2) = \left[\hat{M}_D(x, x+l-1), \hat{M}_D(i+2, j-l)\right], \vec{\hat{M}}(b_3) = \left[\hat{M}_D(2, \Delta L^{D1} + 1)\right]\right\}$$

with $l = 1,..., j - i - 2$ and $x = i + 2,..., j - l + 1$ ($j - i \geq 3$ must hold); (H-2) the case where an insertion and a subsequent "long" deletion occur along b_2, represented as

$$\left\{\vec{\hat{M}}(b_1) = [\ ], \vec{\hat{M}}(b_2) = \left[\hat{M}_I(x,l), \hat{M}_D(i+2, j+l)\right], \vec{\hat{M}}(b_3) = \left[\hat{M}_D(2, \Delta L^{D1} + 1)\right]\right\}$$

with $l = 1,...,\min\{L^{CO}_{I:2}, L^{CO}_{D:2} - j + i + 1\}$ and $x = i + 1,..., j$; (H-3) the case where two successive deletions occur along b_3, represented as

$$\left\{\vec{\hat{M}}(b_1) = [\ ], \vec{\hat{M}}(b_2) = \left[\hat{M}_D(i+2, j)\right], \vec{\hat{M}}(b_3) = \left[\hat{M}_D(x, x+l-1), \hat{M}_D(2, \Delta L^{D1} - l + 1)\right]\right\}$$

with $l = 1,...,\Delta L^{D1} - 1$ and $x = 2,...,\Delta L^{D1} - l + 2$; and (H-4) the case where an insertion and a subsequent "long" deletion occur along b_3, represented as

$$\left\{\vec{\hat{M}}(b_1) = [\ ], \vec{\hat{M}}(b_2) = \left[\hat{M}_D(i+2, j)\right], \vec{\hat{M}}(b_3) = \left[\hat{M}_I(x,l), \hat{M}_D(2, \Delta L^{D1} + l + 1)\right]\right\}$$

with $l = 1,...,\min\{L^{CO}_{I:3}, L^{CO}_{D:3} - \Delta L^{D1}\}$ and $x = 1,...,\Delta L^{D1} + 1$.

Third, type (I) does not need be sub-classified, and each history of this type can be represented as:

$$\left\{\vec{\hat{M}}(b_1) = \left[\hat{M}_I(i'+1, j'-i'-1)\right], \vec{\hat{M}}(b_2) = \left[\hat{M}_D(i+2, i'+j-j'+1)\right], \vec{\hat{M}}(b_3) = \left[\hat{M}_D(2, \Delta L^{D1} - j' + i' + 2)\right]\right\},$$

which is determined uniquely from the root state ($s^{Root} = \left[L, 1,..., i', j',..., \Delta L^{D1}, R\right]$ with $i' \geq i$, $i' + 2 \leq j' \leq j$, but also with $(i', j') \neq (i, j)$). Fourth, type (J) does not need be sub-classified, either, and each history of this type can be represented as:

$$\left\{\vec{\hat{M}}(b_1) = \left[\hat{M}_D(k+2, k+l+1)\right], \vec{\hat{M}}(b_2) = \left[\hat{M}_D(i+2, j+l)\right], \vec{\hat{M}}(b_3) = \left[\hat{M}_D(2, \Delta L^{D1} + l + 1)\right]\right\},$$

which, again, is determined uniquely from the root state ($s^{Root} = \left[L, 1,..., k, \upsilon'_1,..., \upsilon'_l, k+1,..., \Delta L^{D1}, R\right]$).

Thus, the summed contributions from the next-fewest-indel local histories in case (IV) can be expressed as:

$$\tilde{\tilde{\mathrm{M}}}_P\left[\Lambda^{ID}_{\Psi}\left[N_{\mathrm{K}} = 2; C_{\mathrm{K}}; \alpha[s_1, s_2, s_3]; T\right] \middle| T\right]$$
$$= \left(\sum_{\nu=1,2,3,4} \tilde{\tilde{\mathrm{M}}}_P[(G-\nu)]\right) + \left(\sum_{\nu=1,2,3,4} \tilde{\tilde{\mathrm{M}}}_P[(H-\nu)]\right) + \tilde{\tilde{\mathrm{M}}}_P[(I)] + \tilde{\tilde{\mathrm{M}}}_P[(J)].$$

--- Eq.(A2.3)

Here $\tilde{\breve{\mathrm{M}}}_P[(X)]$ (with $X = G-1,...,J$) represents the summed contribution from the histories over one of the aforementioned 10 sub-types. Each such term can be calculated by summing the contributions from individual local histories belonging to each sub-type over the possible lengths and possible positions, if necessary. Under a space-homogeneous model, the summation over possible positions will be reduced to the multiplication by the number of possible positions (for each fixed set of lengths). And the contributions from individual local histories can be calculated according to the definition, Eq.(1.1.2b) supplemented by Eqs.(4.2.4b,6b,8) of part I, and taking advantage of, or by slightly modifying, Eqs.(1.2.4,5d,5e) and Eqs.(A1.1.1,2d,2e) for contributions from single-event and two-event local histories for PWAs. In the following, we will give the expressions for the contributions from individual local histories of respective sub-types under Dawg's indel model.

Type (G-1):

$$\breve{\mathrm{M}}_P\left[\left\{\begin{array}{l}\bar{\hat{M}}(b_1)=\left[\hat{M}_I(i+1,l),\hat{M}_I(x,j-i-1-l)\right],\\ \bar{\hat{M}}(b_2)=[\ \ ],\bar{\hat{M}}(b_3)=\left[\hat{M}_D(2,\Delta L^{D1}-j+i+2)\right]\end{array}\right\}\middle|\left(\begin{array}{l}s^{Root}=s_0^{Root}=s_2\\ =\left[L,1,...,,i,j,...,\Delta L^{D1},R\right],n^{Root}\end{array}\right)\right]$$

$$=\frac{\lambda_{I:1}f_{I:1}(l)\lambda_{I:1}f_{I:1}(j-i-1-l)}{(\lambda_{I:1}+\lambda_{D:1})(j-i-1-l)}\left[\frac{1-e^{-(\lambda_{I:1}+\lambda_{D:1})l(t_{F:1}-t_I)}}{(\lambda_{I:1}+\lambda_{D:1})\,l}-\frac{1-e^{-(\lambda_{I:1}+\lambda_{D:1})(j-i-1)(t_{F:1}-t_I)}}{(\lambda_{I:1}+\lambda_{D:1})(j-i-1)}\right]$$

$$\times\lambda_{D:3}f_{D:3}(\Delta L^{D1}-j+i+1)\frac{\exp\left((\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)\,(t_{F:3}-t_I)\right)-1}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)}\quad .$$

--- Eq.(A2.4a)

Type (G-2):

$$\breve{\mathrm{M}}_P\left[\left\{\begin{array}{l}\bar{\hat{M}}(b_1)=\left[\hat{M}_I(i+1,j-i-1+l),\hat{M}_D(x,x+l-1)\right],\\ \bar{\hat{M}}(b_2)=[\ \ ],\bar{\hat{M}}(b_3)=\left[\hat{M}_D(2,\Delta L^{D1}-j+i+2)\right]\end{array}\right\}\middle|\left(\begin{array}{l}s^{Root}=s_0^{Root}=s_2\\ =\left[L,1,...,,i,j,...,\Delta L^{D1},R\right],n^{Root}\end{array}\right)\right]$$

$$=\frac{\lambda_{I:1}f_{I:1}(j-i-1+l)\lambda_{D:1}f_{D:1}(l)}{(\lambda_{I:1}+\lambda_{D:1})\,l}\left[\frac{1-e^{-(\lambda_{I:1}+\lambda_{D:1})(j-i-1)(t_{F:1}-t_I)}}{(\lambda_{I:1}+\lambda_{D:1})(j-i-1)}-\frac{1-e^{-(\lambda_{I:1}+\lambda_{D:1})(j-i-1+l)(t_{F:1}-t_I)}}{(\lambda_{I:1}+\lambda_{D:1})(j-i-1+l)}\right]$$

$$\times\lambda_{D:3}f_{D:3}(\Delta L^{D1}-j+i+1)\frac{\exp\left((\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)\,(t_{F:3}-t_I)\right)-1}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)}\quad .$$

--- Eq.(A2.4b)

Type (G-3):

$$\breve{\mathbf{M}}_P\left[\left\{\begin{array}{l}\vec{\hat{M}}(b_1)=\left[\hat{M}_I(i+1,\,j-i-1)\right],\,\vec{\hat{M}}(b_2)=[\ \ ],\\ \vec{\hat{M}}(b_3)=\left[\hat{M}_D(x,x+l-1),\,\hat{M}_D(2,\,\Delta L^{D1}-j+i-l+2)\right]\end{array}\right\}\middle|\left(\begin{array}{l}s^{Root}=s_0^{Root}=s_2\\ =\left[L,1,\ldots,,i,\,j,\ldots,\Delta L^{D1},\,R\right],\,n^{Root}\end{array}\right)\right]$$

$$=\lambda_{I:1}f_{I:1}(j-i-1)\,\frac{1-\exp\left(-(\lambda_{I:1}+\lambda_{D:1})(j-i-1)(t_{F:1}-t_I)\right)}{(\lambda_{I:1}+\lambda_{D:1})(j-i-1)}$$

$$\times\frac{\lambda_{D:3}f_{D:3}(l)\,\lambda_{D:3}f_{D:3}(\Delta L^{D1}-j+i+1-l)}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1-l)}\left[\frac{e^{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)(t_{F:3}-t_I)}-1}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)}-\frac{e^{(\lambda_{I:3}+\lambda_{D:3})l(t_{F:3}-t_I)}-1}{(\lambda_{I:3}+\lambda_{D:3})l}\right]\quad.$$

--- Eq.(A2.4c)

Type (G-4):

$$\breve{\mathbf{M}}_P\left[\left\{\begin{array}{l}\vec{\hat{M}}(b_1)=\left[\hat{M}_I(i+1,\,j-i-1)\right],\,\vec{\hat{M}}(b_2)=[\ \ ],\\ \vec{\hat{M}}(b_3)=\left[\hat{M}_I(x,l),\,\hat{M}_D(2,\,\Delta L^{D1}-j+i+l+2)\right]\end{array}\right\}\middle|\left(\begin{array}{l}s^{Root}=s_0^{Root}=s_2\\ =\left[L,1,\ldots,,i,\,j,\ldots,\Delta L^{D1},\,R\right],\,n^{Root}\end{array}\right)\right]$$

$$=\lambda_{I:1}f_{I:1}(j-i-1)\,\frac{1-\exp\left(-(\lambda_{I:1}+\lambda_{D:1})(j-i-1)(t_{F:1}-t_I)\right)}{(\lambda_{I:1}+\lambda_{D:1})(j-i-1)}$$

$$\times\frac{\lambda_{I:3}f_{I:3}(l)\,\lambda_{D:3}f_{D:3}(\Delta L^{D1}-j+i+1+l)}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1+l)}\left[\frac{e^{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)(t_{F:3}-t_I)}-1}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j+i+1)}-\frac{1-e^{-(\lambda_{I:3}+\lambda_{D:3})l(t_{F:3}-t_I)}}{(\lambda_{I:3}+\lambda_{D:3})l}\right]\quad.$$

--- Eq.(A2.4d)

Type (H-1):

$$\breve{\mathbf{M}}_P\left[\left\{\begin{array}{l}\vec{\hat{M}}(b_2)=\left[\hat{M}_D(x,\,x+l-1),\,\hat{M}_D(i+2,\,j-l)\right],\\ \vec{\hat{M}}(b_1)=[\ \ ],\,\vec{\hat{M}}(b_3)=\left[\hat{M}_D(2,\,\Delta L^{D1}+1)\right]\end{array}\right\}\middle|\left(\begin{array}{l}s^{Root}=s_1\\ =\left[L,1,\ldots,\Delta L^{D1},\,R\right],\,n^{Root}\end{array}\right)\right]$$

$$=\exp\left(-\sum_{m=1,2,3}(\lambda_{I:m}+\lambda_{D:m})(j-i-1)(t_{F:m}-t_I)\right)$$

$$\times\frac{\lambda_{D:2}f_{D:2}(l)\,\lambda_{D:2}f_{D:2}(j-i-1-l)}{(\lambda_{I:2}+\lambda_{D:2})(j-i-1-l)}\left[\frac{e^{(\lambda_{I:2}+\lambda_{D:2})(j-i-1)(t_{F:2}-t_I)}-1}{(\lambda_{I:2}+\lambda_{D:2})(j-i-1)}-\frac{e^{(\lambda_{I:2}+\lambda_{D:2})l(t_{F:2}-t_I)}-1}{(\lambda_{I:2}+\lambda_{D:2})l}\right]$$

$$\times\lambda_{D:3}f_{D:3}(\Delta L^{D1})\,\frac{\exp\left(+(\lambda_{I:3}+\lambda_{D:3})\Delta L^{D1}(t_{F:3}-t_I)\right)-1}{(\lambda_{I:3}+\lambda_{D:3})\Delta L^{D1}}\quad.$$

--- Eq.(A2.4e)

Type (H-2):

$$\breve{\mathrm{M}}_P\left[\left\{\begin{array}{l}\vec{\hat{M}}(b_2)=\left[\hat{M}_I(x,l),\hat{M}_D(i+2,j+l)\right],\\ \vec{\hat{M}}(b_1)=[\ \ ],\vec{\hat{M}}(b_3)=\left[\hat{M}_D(2,\Delta L^{D1}+1)\right]\end{array}\right\}\middle|\left(\begin{array}{l}s^{Root}=s_1\\ =\left[L,1,\ldots,\Delta L^{D1},R\right],n^{Root}\end{array}\right)\right]$$

$$=\exp\left(-\sum_{m=1,2,3}(\lambda_{I:m}+\lambda_{D:m})(j-i-1)(t_{F:m}-t_I)\right)$$

$$\times\frac{\lambda_{I:2}f_{I:2}(l)\,\lambda_{D:2}f_{D:2}(j-i-1+l)}{(\lambda_{I:2}+\lambda_{D:2})(j-i-1+l)}\left[\frac{e^{(\lambda_{I:2}+\lambda_{D:2})(j-i-1)(t_{F:2}-t_I)}-1}{(\lambda_{I:2}+\lambda_{D:2})(j-i-1)}-\frac{1-e^{-(\lambda_{I:2}+\lambda_{D:2})l(t_{F:2}-t_I)}}{(\lambda_{I:2}+\lambda_{D:2})l}\right]$$

$$\times\lambda_{D:3}f_{D:3}(\Delta L^{D1})\frac{\exp\left(+(\lambda_{I:3}+\lambda_{D:3})\Delta L^{D1}(t_{F:3}-t_I)\right)-1}{(\lambda_{I:3}+\lambda_{D:3})\Delta L^{D1}}\quad.$$

--- Eq.(A2.4f)

Type (H-3):

$$\breve{\mathrm{M}}_P\left[\left\{\begin{array}{l}\vec{\hat{M}}(b_1)=[\ \ ],\vec{\hat{M}}(b_2)=\left[\hat{M}_D(i+2,j)\right],\\ \vec{\hat{M}}(b_3)=\left[\hat{M}_D(x,x+l-1),\hat{M}_D(2,\Delta L^{D1}-l+1)\right]\end{array}\right\}\middle|\left(\begin{array}{l}s^{Root}=s_1\\ =\left[L,1,\ldots,\Delta L^{D1},R\right],n^{Root}\end{array}\right)\right]$$

$$=\exp\left(-\sum_{m=1,2,3}(\lambda_{I:m}+\lambda_{D:m})(j-i-1)(t_{F:m}-t_I)\right)$$

$$\times\lambda_{D:2}f_{D:2}(j-i-1)\frac{\exp\left(+(\lambda_{I:2}+\lambda_{D:2})(j-i-1)(t_{F:2}-t_I)\right)-1}{(\lambda_{I:2}+\lambda_{D:2})(j-i-1)}$$

$$\times\frac{\lambda_{D:3}f_{D:3}(l)\,\lambda_{D:3}f_{D:3}(\Delta L^{D1}-l)}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-l)}\left[\frac{e^{(\lambda_{I:3}+\lambda_{D:3})\Delta L^{D1}(t_{F:3}-t_I)}-1}{(\lambda_{I:3}+\lambda_{D:3})\Delta L^{D1}}-\frac{e^{(\lambda_{I:3}+\lambda_{D:3})l(t_{F:3}-t_I)}-1}{(\lambda_{I:3}+\lambda_{D:3})l}\right]\quad.$$

--- Eq.(A2.4g)

Type (H-4):

$$\breve{\mathrm{M}}_P\left[\left\{\begin{array}{l}\vec{\hat{M}}(b_1)=[\ \ ],\vec{\hat{M}}(b_2)=\left[\hat{M}_D(i+2,j)\right],\\ \vec{\hat{M}}(b_3)=\left[\hat{M}_I(x,l),\hat{M}_D(2,\Delta L^{D1}+l+1)\right]\end{array}\right\}\middle|\left(\begin{array}{l}s^{Root}=s_1\\ =\left[L,1,\ldots,\Delta L^{D1},R\right],n^{Root}\end{array}\right)\right]$$

$$=\exp\left(-\sum_{m=1,2,3}(\lambda_{I:m}+\lambda_{D:m})(j-i-1)(t_{F:m}-t_I)\right)$$

$$\times\lambda_{D:2}f_{D:2}(j-i-1)\frac{\exp\left(+(\lambda_{I:2}+\lambda_{D:2})(j-i-1)(t_{F:2}-t_I)\right)-1}{(\lambda_{I:2}+\lambda_{D:2})(j-i-1)}$$

$$\times\frac{\lambda_{I:3}f_{I:3}(l)\,\lambda_{D:3}f_{D:3}(\Delta L^{D1}+l)}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}+l)}\left[\frac{e^{(\lambda_{I:3}+\lambda_{D:3})\Delta L^{D1}(t_{F:3}-t_I)}-1}{(\lambda_{I:3}+\lambda_{D:3})\Delta L^{D1}}-\frac{1-e^{-(\lambda_{I:3}+\lambda_{D:3})l(t_{F:3}-t_I)}}{(\lambda_{I:3}+\lambda_{D:3})l}\right]\quad.$$

--- Eq.(A2.4h)

Type (I):

$$\breve{\mathrm{M}}_P\left[\left\{\begin{array}{l}\bar{\hat{M}}(b_1)=\left[\hat{M}_I(i'+1,\,j'-i'-1)\right],\\ \bar{\hat{M}}(b_2)=\left[\hat{M}_D(i+2,\,i'+j-j'+1)\right],\\ \bar{\hat{M}}(b_3)=\left[\hat{M}_D(2,\,\Delta L^{D1}-j'+i'+2)\right]\end{array}\right\}\middle|\begin{pmatrix}s^{Root}\\ =\left[L,1,\ldots,i',\,j',\ldots,\Delta L^{D1},\,R\right],\,n^{Root}\end{pmatrix}\right]$$

$$=\exp\left(-\sum_{m=1,2,3}(\lambda_{I:m}+\lambda_{D:m})(i'-i+j-j')(t_{F:m}-t_I)\right)$$

$$\times\,\lambda_{I:1}f_{I:1}(j'-i'-1)\,\frac{1-\exp\left(-(\lambda_{I:1}+\lambda_{D:1})(j'-i'-1)(t_{F:1}-t_I)\right)}{(\lambda_{I:1}+\lambda_{D:1})(j'-i'-1)}$$

$$\times\,\lambda_{D:2}f_{D:2}(i'-i+j-j')\,\frac{\exp\left(+(\lambda_{I:2}+\lambda_{D:2})(i'-i+j-j')(t_{F:2}-t_I)\right)-1}{(\lambda_{I:2}+\lambda_{D:2})(i'-i+j-j')}$$

$$\times\,\lambda_{D:3}f_{D:3}(\Delta L^{D1}-j'+i'+1)\,\frac{\exp\left(+(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j'+i'+1)(t_{F:3}-t_I)\right)-1}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}-j'+i'+1)}\,.$$

--- Eq.(A2.4i)

It should be noted that the right hand side of this equation depends on i' and j' only through $l'\equiv j'-i'-1$. Thus, it follows that there are $j-i-l'$ histories with the same l' (and therefore the same probability) but with different i''s (namely, $i'=i,\ldots,j-1-l'$).

Type (J):

$$\breve{\mathrm{M}}_P\left[\left\{\begin{array}{l}\bar{\hat{M}}(b_1)=\left[\hat{M}_D(k+2,\,k+l+1)\right],\\ \bar{\hat{M}}(b_2)=\left[\hat{M}_D(i+2,\,j+l)\right],\\ \bar{\hat{M}}(b_3)=\left[\hat{M}_D(2,\,\Delta L^{D1}+l+1)\right]\end{array}\right\}\middle|\begin{pmatrix}s^{Root}\\ =\left[L,1,\ldots,k,\,\upsilon_1',\ldots,\upsilon_l',\,k+1,\ldots,\Delta L^{D1},\,R\right],\,n^{Root}\end{pmatrix}\right]$$

$$=\exp\left(-\sum_{m=1,2,3}(\lambda_{I:m}+\lambda_{D:m})(j-i-1+l)(t_{F:m}-t_I)\right)\times\lambda_{D:1}f_{D:1}(l)\,\frac{\exp\left(+(\lambda_{I:1}+\lambda_{D:1})\,l\,(t_{F:1}-t_I)\right)-1}{(\lambda_{I:1}+\lambda_{D:1})\,l}$$

$$\times\,\lambda_{D:2}f_{D:2}(j-i-1+l)\,\frac{\exp\left(+(\lambda_{I:2}+\lambda_{D:2})(j-i-1+l)(t_{F:2}-t_I)\right)-1}{(\lambda_{I:2}+\lambda_{D:2})(j-i-1+l)}$$

$$\times\,\lambda_{D:3}f_{D:3}(\Delta L^{D1}+l)\,\frac{\exp\left(+(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}+l)(t_{F:3}-t_I)\right)-1}{(\lambda_{I:3}+\lambda_{D:3})(\Delta L^{D1}+l)}\,.$$

--- Eq.(A2.4j)

It should be noted that the right hand side of this equation does not depend on k whereas it does depend on l. Thus, it follows that there are $j-i$ histories with the same l (and therefore the same probability) but with different k's (namely, $k=i,\ldots,j-1$).

Summing each of Eqs.(A2.4a-j) over the lengths and positions specified above Eq.(A2.3), we obtain $\tilde{\tilde{\mathrm{M}}}_P[(X)]$ with $X=G-1,\ldots,J$, the summation of which

(according to Eq.(A2.3)) gives the total contribution from the next-fewest-indel local histories.

Tables 1-4

Table 1. Various "threshold gap lengths" in case (ii) of local PWAs

$X = (\lambda_I + \lambda_D)(t_F - t_I)$ [a]	$(\Delta L^A)_{0.5}^{(NP-ii)}$ [b]	$(\Delta L^A)_{0.5}^{(1-ii)}$ [c]	$(\Delta L^A)_{0.5}^{(2-ii)}$ [c]	$(\Delta L^A)_{0.5}^{(5-ii)}$ [c]
0.01 indels/site	128	160	> 300	> 300
0.04 indels/site	31	41	99	272
0.1 indels/site	12	17	42	119
0.2 indels/site	6	8	22	66
Approximate relation [d]	$Y \approx 1.2/X$	$Y \approx 1.6/X$	$Y \sim 4/X$	$Y \sim 11/X$ (?)

NOTE: The parameters used for this analysis are: $\lambda_I = \lambda_D = 0.1$, $L_I^{CO} = L_D^{CO} = 500$, and $f_I(l) = f_D(l) = l^{-1.6} \Big/ \left(\sum_{k=1}^{500} k^{-1.6} \right)$. For the iteration analysis, we used $\Delta L_{\max}^A = \Delta L_{\max}^D = 300$ and $(\lambda_I + \lambda_D)(t_F - t_I)/N_P = 0.002$ as well. Because of the symmetry under the time reversal, the identical results apply also to the local PWAs in case (iii), if we replace $(\Delta L^A)_{0.5}^{(xx-ii)}$ with $(\Delta L^D)_{0.5}^{(xx-iii)}$.

[a] The expected number of indels per site.
[b] The number of ancestral sites in between the PASs, *i.e.,* ΔL^A, at which the total probability of the next-parsimonious local indel histories is 1/2 (=0.5) of that of the parsimonious local indel histories.
[c] $(\Delta L^A)_{0.5}^{(N_{ID}-ii)}$ is the value of ΔL^A at which the local histories involving up to (and including) N_{ID} indels each account for 1/2 (=0.5) of the total probability of all local histories consistent with the configuration. We used the probability calculated up to (and including) $N_{ID} = 50$ indel steps as a proxy of the total probability.
[d] A rough (inversely proportional) relationship between each threshold gap length (Y) and the expected number of indels per site (X).

Table 2. Perturbation analysis on local PWA probabilities in case (iv)

$(\Delta L^A, \Delta L^D)$	0.01 indels/site	0.04 indels/site	0.1 indels/site	0.2 indels/site
(1, 1)	**0.003**	**0.010**	**0.024**	**0.045**
(3, 1)	**0.021**	**0.084**	**0.204**	**0.393**
(3, 3)	**0.042**	**0.166**	**0.402**	0.768
(5, 5)	**0.073**	**0.283**	0.672	1.256
(10, 1)	**0.064**	**0.246**	0.572	1.013
(10, 10)	**0.149**	0.561	1.292	2.288
(25, 1)	**0.151**	0.547	1.112	1.541
(25, 4)	**0.198**	0.723	1.519	2.234
(30, 10)	**0.288**	1.038	2.164	3.072
(100, 1)	0.537	1.333	1.507	1.574
(100, 3)	0.607	1.593	1.894	2.033
(300, 1)	1.165	1.394	1.427	1.527

NOTE: Each cell shows the ratio of the total probability of the next-fewest indel histories to that of the fewest indel histories, when there are ΔL^A ancestral sites and ΔL^D descendant sites in between the PASs. Each column is labeled with the expected number of indels per site (*i.e.*, $(\lambda_I + \lambda_D)(t_F - t_I)$). The parameters used for this analysis are: $\lambda_I = \lambda_D = 0.1$, $L_I^{CO} = L_D^{CO} = 500$, and $f_I(l) = f_D(l) = l^{-1.6} \Big/ \left(\sum_{k=1}^{500} k^{-1.6} \right)$. Because of the symmetry of the probabilities under the time reversal, the ratio for $(\Delta L^A, \Delta L^D) = (L_1, L_2)$ is identical to that for $(\Delta L^A, \Delta L^D) = (L_2, L_1)$. Thus we only showed the results for $\Delta L^A \geq \Delta L^D$. The ratios that are less than 0.5 are shown in boldface.

Table 3. Perturbation analysis on local MSA probabilities in case (IV)

$(\Delta L^{D1}, \Delta L^{D2})$	0.01 indels/site	0.04 indels/site	0.1 indels/site	0.2 indels/site
(2, 1)	**0.004**	**0.016**	**0.037**	**0.067**
(3, 1)	**0.016**	**0.063**	**0.150**	**0.279**
(3, 2)	**0.012**	**0.049**	**0.118**	**0.225**
(10, 1)	**0.050**	**0.190**	**0.432**	0.751
(10, 2)	**0.060**	**0.225**	0.509	0.895
(10, 8)	**0.060**	**0.231**	0.543	0.980
(10, 9)	**0.047**	**0.182**	**0.421**	0.740
(30, 1)	**0.140**	**0.487**	0.915	1.203
(30, 5)	**0.163**	0.563	1.118	1.645
(30, 25)	**0.173**	0.620	1.251	1.850
(30, 29)	**0.139**	**0.486**	0.909	1.189
(100, 1)	**0.418**	0.985	1.180	1.304
(100, 99)	**0.418**	0.981	1.170	1.290

NOTE: Each cell shows the ratio of the total probability of the next-fewest indel histories to that of the fewest indel histories when sequence 1 and sequence 2 have ΔL^{D1} sites and $\Delta L^{D2} \equiv \Delta L^{D1} - (j - i - 1)$ ($< \Delta L^{D1}$) sites, respectively, in between the PASs. Each column is labeled with the expected number of indels per site (*i.e.*, $(\lambda_{I:m} + \lambda_{D:m})(t_{F:m} - t_I)$, which is independent of m ($\in \{1, 2, 3\}$) in this case). The parameters used for this analysis are: $\lambda_{I:m} = \lambda_{D:m} = 0.1$, $L^{CO}_{I:m} = L^{CO}_{D:m} = 500$ and $f_{I:m}(l) = f_{D:m}(l) = l^{-1.6} \Big/ \left(\sum_{k=1}^{500} k^{-1.6} \right)$, for $^{\forall} m = 1, 2, 3$. And we used the 3-OTU trees with $t_{F:1} = t_{F:2} = t_{F:3}$. The ratios that are less than 0.5 are shown in boldface.

Table 4. Comparison of Kim-Sinha's probabilities of case (iv) local PWAs with reference probabilities

$(\Delta L^A, \Delta L^D)$	Ratio (Eq.(VII-2.13))	Overlapping [a]
(1, 1)	1.667	0.167
(3, 1)	1.883	0.383
(3, 3)	2.449	0.949
(5, 5)	3.325	1.825
(10, 1)	2.165	0.665
(10, 10)	5.572	4.072
(25, 1)	2.355	0.855
(25, 4)	4.714	3.214
(30, 10)	8.300	6.800
(100, 1)	2.561	1.061
(100, 3)	4.896	3.396
(300, 1)	2.659	1.159

NOTE: Each cell in the middle column gives the ratio, Eq.(2.2.13), of the reference probability in the fewest-indel approximation to the corresponding probability given by Kim and Sinha's HMM (2007), for a local PWA with ΔL^A ancestral sites and ΔL^D descendant sites in between a pair of PASs. The parameters used for this analysis are: $\lambda_I = \lambda_D = 0.1$, $L_I^{CO} = L_D^{CO} = 5000$, and $f_I(l) = f_D(l) = l^{-1.6} \Big/ \left(\sum_{k=1}^{5000} k^{-1.6} \right)$. The ratio for $(\Delta L^A, \Delta L^D) = (L_1, L_2)$ is identical to that for $(\Delta L^A, \Delta L^D) = (L_2, L_1)$. Thus, we only showed the results for $\Delta L^A \geq \Delta L^D$.

[a] The effect of overlapping indels, which is the second term on the right hand side of Eq.(2.2.13).

Figures 1-7 (with legends)

A. Original local MSA (in S^{II})

R		…	3	4	5	-	-	6	…
5		…	3	4	5	-	E	6	…
6		…	3	4	5	-	-	6	…
1		…	3	-	-	-	E	6	…
2		…	3	4	5	F	E	6	…
3		…	3	4	5	-	-	6	…
4		…	3	-	5	-	-	6	…

C_4

B. Reassigning ancestries

R		…	L	1	2	-	-	R	…
5		…	L	1	2	-	3	R	…
6		…	L	1	2	-	-	R	…
1		…	L	-	-	-	3	R	…
2		…	L	1	2	4	3	R	…
3		…	L	1	2	-	-	R	…
4		…	L	-	2	-	-	R	…

C_4

C. Reassigned site-numbers and $\Delta L(s)$

$s^{Root} = s_6 = s_3$:

1	2	3	4
L	1	2	R

$\Delta L(s^{Root}) = \Delta L(s_6) = \Delta L(s_3) = 2$

s_5 :

1	2	3	4	5
L	1	2	3	R

$\Delta L(s_5) = 3$

s_1 :

1	2	3
L	3	R

$\Delta L(s_1) = 1$

s_2 :

1	2	3	4	5	6
L	1	2	4	3	R

$\Delta L(s_2) = 4$

s_4 :

1	2	3
L	2	R

$\Delta L(s_4) = 1$

Figure 1. Typical situation and notation considered in Section 1 of Results.
A. As an illustration, we use the local MSA confined in the region C_4 of the MSA in the Figure 11 B of part I (Ezawa, Graur and Landan 2015a), and re-assign the ancestries as shown in panel **B**. The ancestries in between the PASs (with the re-assigned ancestries L and R) are just an example and not always assigned in this way. **C.** Sequences extracted from the MSA. Shown above each site is the site number (*i.e.*, spatial coordinate) assigned to it. And shown on the right of each sequence is the count of sites in between the PASs.
In panels A and B, as in Figure 5 of part I, the boldface characters in the leftmost columns stand for the sequence IDs. (More precisely, the number 'i' stands for the sequence s_i, and the 'R' stands for the root sequence (s^{Root}).)

A. Case (i)

A		L	R
D		L	R

B. Case (ii)

A		L	1	2	3	R
D		L	-	-	-	R

C. Case (iii)

A		L	-	-	R
D		L	υ_1	υ_2	R

D. Case (iv)

A		L	1	2	3	-	-	R
D		L	-	-	-	$\upsilon^D{}_1$	$\upsilon^D{}_2$	R

or

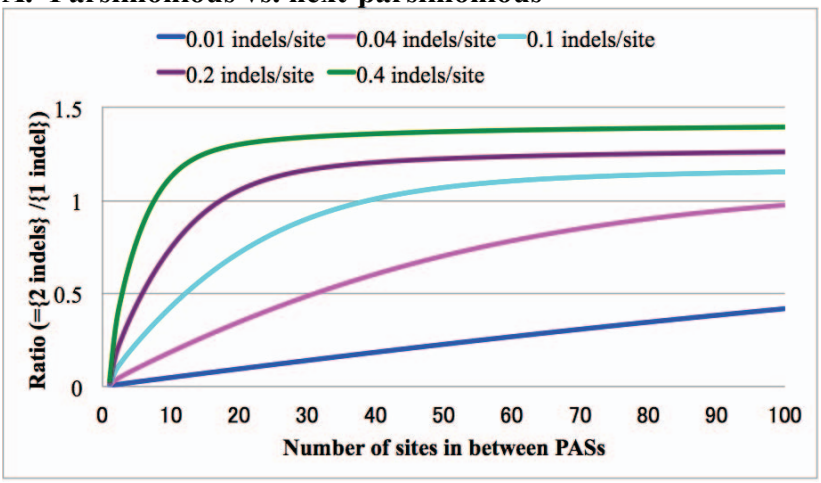
A		L	-	-	1	2	3	R
D		L	$\upsilon^D{}_1$	$\upsilon^D{}_2$	-	-	-	R

Figure 2. Examples of four types of local gap configurations in PWA between ancestral and descendant sequences.

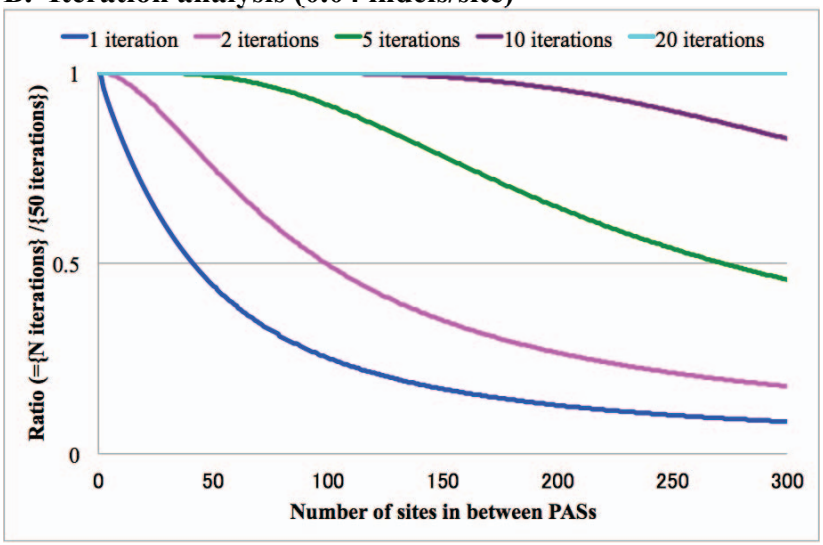
A. Case (i). **B.** Case (ii) with $\Delta L^A = 3$. **C.** Case (iii) with $\Delta L^D = 2$. **D**. Case (iv) with $\Delta L^A = 3$ and $\Delta L^D = 2$.

In each PWA, each site (a cell) is assigned an ancestry. In the leftmost column of each PWA, the boldface italic 'A' and 'D' stand for an ancestor (s^A) and a descendant (s^D), respectively. The magenta boxes and the cyan boxes represent unpreserved ancestral sites and inserted descendant sites, respectively. In panel D, the PWA on the right (in parentheses) is equivalent to the PWA on the left, as far as the homology structure alone is concerned.

A. Parsimonious vs. next-parsimonious



B. Iteration analysis (0.04 indels/site)



C. Iteration analysis (0.2 indels/site)

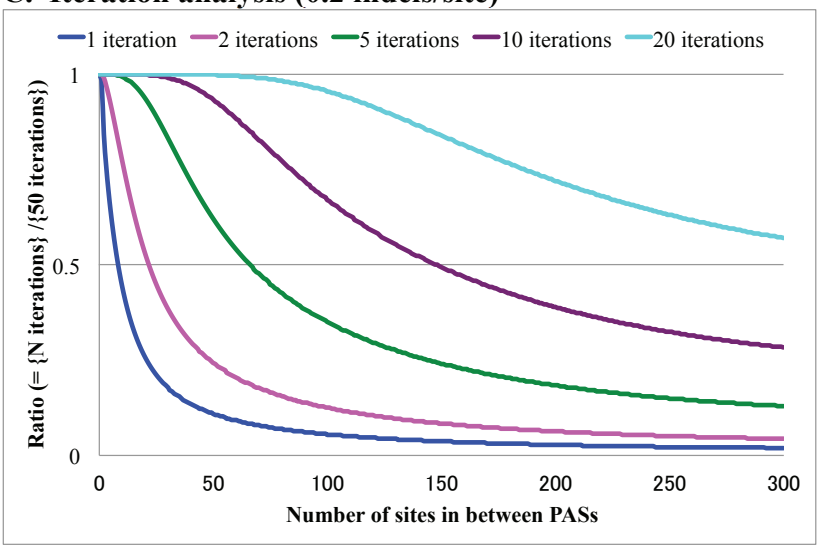
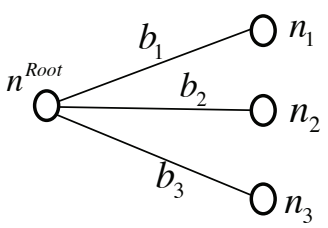


Figure 3. Perturbation analyses on local PWA probabilities.
The results shown in all panels of this figure apply to both case (ii) and case (iii) local PWAs. In all panels, the abscissa is ΔL^A in case (ii) and ΔL^D in case (iii). Panel **A** shows the ratio of the total probability of the next-fewest-indel histories to that of the

fewest-indel history. The blue, magenta, cyan, purple and green curves show the ratios when the expected numbers of indels per site, *i.e.*, the values of $(\lambda_I + \lambda_D)(t_F - t_I)$, are 0.01, 0.04, 0.1, 0.2 and 0.4, respectively. Panels **B** and **C** show the ratios of the multiplication factors calculated *via* the iteration formulas (either Eq.(1.2.8) or Eq.(A1.3.3) in Appendix A1), including histories with up to $N_{ID} = 1$ (blue), 2 (magenta), 5 (green), 10 (purple), and 20 (cyan) indels, to the factor including histories with up to $N_{ID} = 50$ indels. Panel B is for $(\lambda_I + \lambda_D)(t_F - t_I) = 0.04$ indels per site, and panel C is for $(\lambda_I + \lambda_D)(t_F - t_I) = 0.2$ indels per site. The following parameters were used for all panels: $f_I(l) = f_D(l) \propto l^{-1.6}$, $L_I^{CO} = L_D^{CO} = 500$ and $\lambda_I / \lambda_D = 1$. For panels B and C, we also set the following parameters: $\Delta L_{\max}^A = \Delta L_{\max}^D = 300$ and $(\lambda_I + \lambda_D)(t_F - t_I)/N_P = 0.002$.

A. Phylogenetic tree



B. Case (I)

1		L	R
2		L	R
3		L	R

C. Case (II)

1		L	1	2	3	4	5	R
2		L	1	2	3	4	5	R
3		L	-	-	-	-	-	R

D. Case (III)

1		L	1	2	3	4	5	R
2		L	-	-	-	-	-	R
3		L	-	-	-	-	-	R

E. Case (IV)

1		L	1	2	3	4	5	R
2		L	1	2	-	-	5	R
3		L	-	-	-	-	-	R

Figure 4. Typical local gap configurations in MSAs considered for perturbation analyses in Subsection 1.3 of Results.
A. The 3-OTU tree used in the perturbation analyses and the notation. A node (open circle) is labeled n_i (external) or n^{Root} (root). A branch is labeled b_i. **B.** Local gap configuration in case (I). **C.** Case (II) with $\Delta L^{D12} = 5$. **D.** Case (III) with $\Delta L^{D1} = 5$. **E.** Case (IV) with $\Delta L^{D1} = 5$, $i = 2$ and $j = 5$.
In this figure, none of the sites are colored, to take account of the possibilities of their preservations, insertions and deletions.

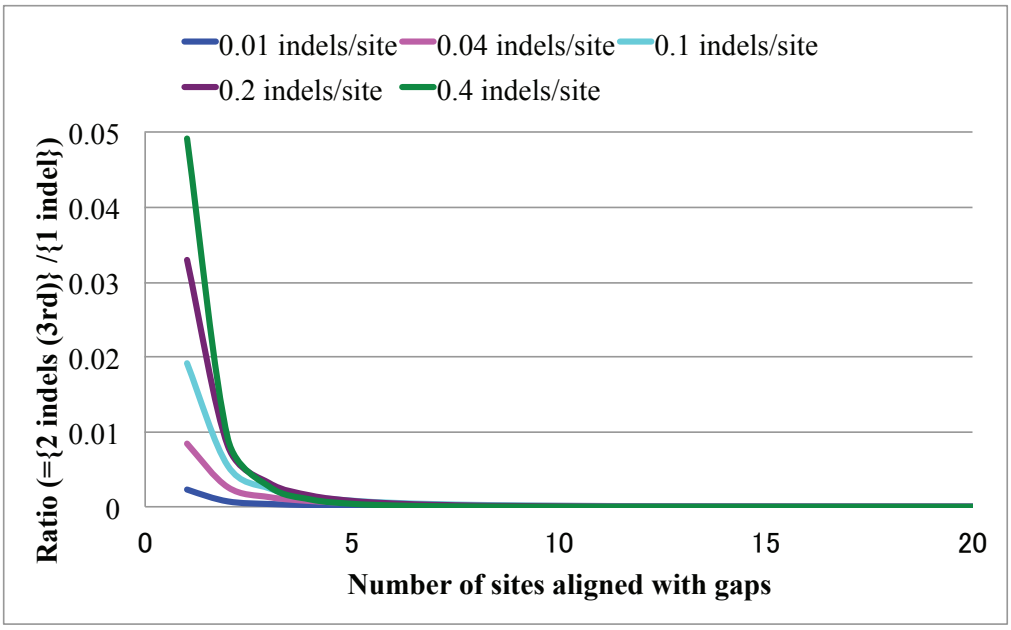
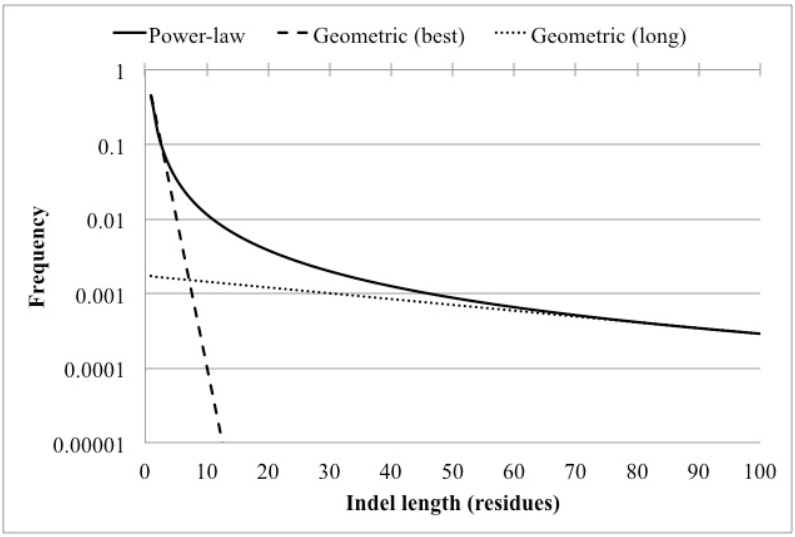


Figure 5. Perturbation analysis on local MSA probabilities in case (III).

The graph shows the ratio of the probability of the history with two deletions giving rise to a case (III) local MSA (Eq.(1.3.13')) to that of the single-insertion history (Eq.(1.3.10)). The ratio is shown as a function of the length of the gapped segment (ΔL^{D1}) and the expected number of indels per site along each branch ($(\lambda_{I:m}+\lambda_{D:m})(t_{F:m}-t_I)$, identical for all $m=1,2,3$). The blue, magenta, cyan, purple, and green curves show the ratios when the values of $(\lambda_{I:m}+\lambda_{D:m})(t_{F:m}-t_I)$ are 0.01, 0.04, 0.1, 0.2 and 0.4 indels/site, respectively. We used a 3-OTU tree whose three branches are equally long. For each branch, we used the same parameters as those used for Figure 3.

A. Distributions



B. Ratio

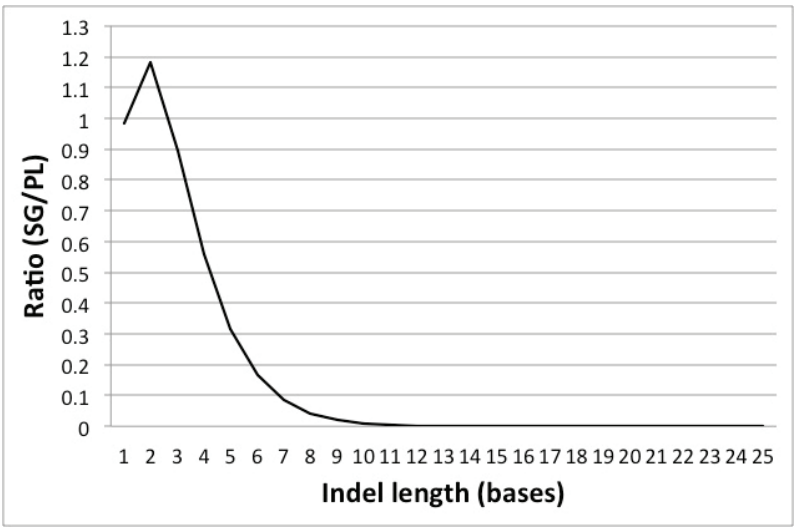


Figure 6. Power-law indel length distribution and its approximation by geometric distributions.

A. Visual comparison of the distributions. The continuous line is the power-law distribution, Eq.(2.1.1) with $a = 1.6$. The dashed line is the scaled geometric distribution that fits the power-law the best according to the least-square criterion (*i.e.*, Eq.(2.1.2) with $A = A_{LS}(1.6) \equiv 0.7125$ and $q = q_{LS}(1.6) \equiv 0.3957$). The dotted line is the scaled geometric distribution that well approximates the behavior of the power-law around 100 residues (*i.e.*, Eq.(2.1.2) with $A = 0.09538$ and $q = 0.9822$). Note the logarithmic scale for the ordinate. **B.** The ratio of the best-fit scaled geometric distribution (SG) to the power-law distribution (PL).

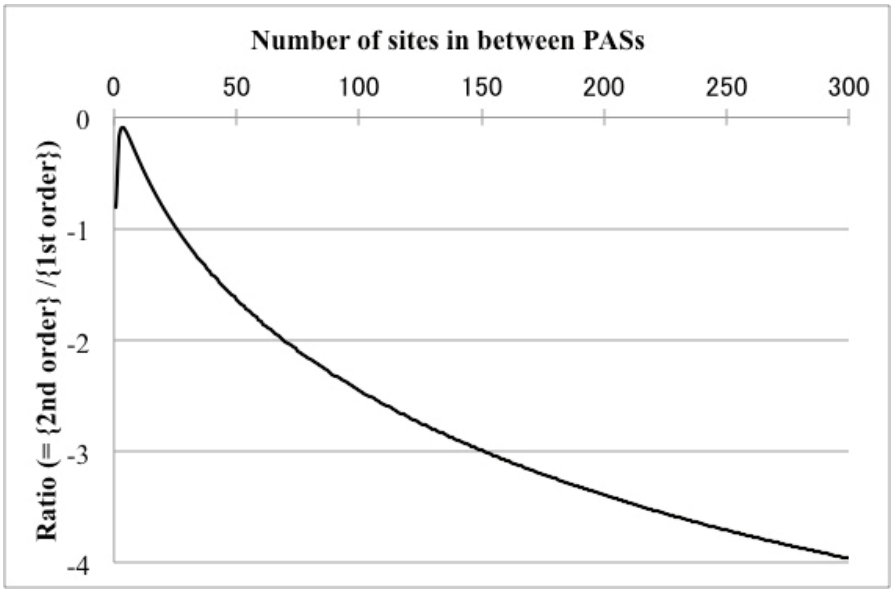


Figure 7. Coefficient of 2nd-order term of PWA probability in cases (ii) and (iii). The curve shows the ratio $G_D(\Delta L^A)/f_D(\Delta L^A)$ as a function of the number of sites in between the PASs (ΔL^A) in a local PWA belonging to case (ii). Here $G_D(\Delta L^A)$ is the coefficient of the 2nd-order term in the expected number of indels ($\lambda|b|$) given in Eq.(2.2.6'b) in Results. And $f_D(\Delta L^A)$ is the coefficient of the 1st-order term (see Eq.(2.2.6'a)). The ratio was calculated with the following parameters: $\lambda_I = \lambda_D = 0.1$, $L_I^{CO} = L_D^{CO} = 5000$, $f_I(l) = f_D(l) = l^{-1.6} \Big/ \left(\sum_{k=1}^{5000} k^{-1.6}\right)$. Under the same parameter setting, this graph also gives the ratio for a case (iii) local PWA as a function of ΔL^D ($G_I(\Delta L^D)/f_I(\Delta L^D)$. See Eqs.(2.2.8a,b)). The value, $\left|1 + \frac{G_D(\Delta L^A)}{f_D(\Delta L^A)}\right|^{-1}$, gives a rough estimate of the "threshold" value of $(\lambda_I + \lambda_D)|b|$ beyond which the violation of the Chapman-Kolmogorov equation (Eq.(3.1.1.1) in part I) may severely undermine the HMM of Kim and Sinha (2007).

References

Benner SA, Cohen MA, Connet GH. 1993. **Empirical and structural models for insertions/deletions as the major path to genomic divergence.** *J Mol Biol.* **229**:1065-1082.

Bishop MJ, Thompson EA. 1986. **Maximum likelihood alignment of DNA sequences.** *J. Mol Biol.* **190**:159-165.

Bradley RK, Holmes I. 2007. **Transducers: an emerging probabilistic framework for modeling indels on trees.** *Bioinformatics* **23**:3258-3262.

Britten RJ. 2002. **Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels.** *Proc. Natl. Acad. Sci. USA* **99**:13633-13635.

Britten RJ, Rowen L, Willians J, Cameron RA. 2003. **Majority of divergence between closely related DNA samples is due to indels.** *Proc. Natl. Acad. Sci. USA* **100**:4661-4665.

Cartwright RA. 2005. **DNA assembly with gap (Dawg): simulating sequence evolution.** *Bioinformatics* **21**:iii31-iii38.

Cartwright RA. 2009. **Problems and solutions for estimating indel rates and length distribution.** *Mol Biol Evol.* **26**:473-480.

Chang MSS, Benner SA. 2004. **Empirical analysis of protein insertions and deletions determining parameters for the correct placement of gaps in protein sequence alignments.** *J Mol Biol.* **341**:617-631.

Chindelevitch L. Li Z, Blais E, Blanchette M. 2006. **On the inference of parsimonious evolutionary scenarios.** *J Bioinform Comput Biol.* **4**:721-744.

Dirac PAM. 1958. *The Principles of Quantum Mechanics, 4th edition*. London, Oxford University Press.

Ezawa K. 2013. LOLIPOG: Log-likelihood for the pattern of gaps in MSA. [http:///www.bioinformatics.org/ftp/pub/lolipog/]

Ezawa K, Graur D, Landan G. 2015a. **Perturbative formulation of general continuous-time Markov model of sequence evolution via insertions/deletions, Part I: Theoretical basis.** (unpublished, to be posted on bioRxiv.)

Ezawa K, Graur D, Landan G. 2015b. **Perturbative formulation of general continuous-time Markov model of sequence evolution via insertions/deletions, Part III: Algorithm for first approximation.** (unpublished, to be posted on bioRxiv.)

Ezawa K, Graur D, Landan G. 2015c. **Perturbative formulation of general continuous-time Markov model of sequence evolution via insertions/deletions, Part IV: Incorporation of substitutions and other mutations.** (unpublished, to be posted on bioRxiv.)

Fan Y, Wang W, Ma G, Liang L, Shi Q, Tao S. 2007. **Patterns of insertion and deletion in mammalian genomes.** *Curr Genomics* **8**:370-378.

Felsenstein J. 1981. **Evolutionary trees from DNA sequences: a maximum likelihood approach.** *J Mol Evol.* **17**:368-376.

Felsenstein J. 2004. *Inferring Phylogenies.* Sunderland (MA), Sinauer Associates.

Gascuel O (editor). 2005. *Mathematics of Evolution and Phylogeny.* New York, Oxford University Press.

Gillespie DT. 1977. **Exact stochastic simulation of coupled chemical reactions.** *J Phys Chem.* **81**:2340-2361.

Gonnet GH, Cohen MA, Benner SA. 1992. **Exhaustive matching of the entire protein sequence database.** *Science* **256**:1443-1445.

Graur D, Li WH. 2000. *Fundamentals of Molecular Evolution, 2nd ed.* Sunderland (MA), Sinauer Associates.

Gu X, Li WH. 1995. **The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment.** *J Mol Evol.* **40**:464-473.

Kent WJ, Baertsch R, Hinrichs A, Miller W, and Haussler D. 2003. **Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes.** *Proc Natl Acad Sci USA* **100**:11484-11489.

Kim J, Sinha S. 2007. **Indelign: a probabilistic framework for annotation of insertions and deletions in a multiple alignment.** *Bioinformatics* **23**:289-297.

Knudsen B, Miyamoto MM. 2003. **Sequence alignments and pair hidden Markov models using evolutionary history.** *J Mol Biol*. **333**:453-460.

Lunter G. 2007. **Probabilistic whole-genome alignments reveal high indel rates in the human and mouse genomes.** *Bioinformatics* **23**:i289-i296.

Lunter GA, Miklós I, Drummond A, Jensen JL, Hein J. 2005. **Bayesian coestimation of phylogeny and sequence alignment.** *BMC Bioinformatics* **6**:83.

Lynch M. 2007. *The Origins of Genome Architecture.* Sunderland (MA), Sinauer Associates.

Metzler D. 2003. **Statistical alignment based on fragment insertion and deletion models.** *Bioinformatics* **19**:490-499.

Messiah A. 1961a. *Quantum Mechanics, Volume 1.* (Translated from French to English by Temmer GM). Amsterdam, North-Holland.

Messiah A. 1961b. *Quantum Mechanics, Volume II.* (Translated from French to English by Potter J). Amsterdam, North-Holland.

Miklós I, Lunter GA, Holmes I. 2004. **A "long indel" model for evolutionary sequence alignment.** *Mol Biol Evol.* **21**:529-540.

Miklós I, Novák Á, Satija R, Lyngsø R, Hein J. 2009. **Stochastic models of sequence evolution including insertion-deletion events.** *Stat Methods Med Res.* **18**:453-485.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 1992. *Numerical Recipes in C: The Art of Scientific Computing, 2nd edition.* Cambridge (UK), Cambridge University Press.

Rivas E. 2005. **Evolutionary models for insertions and deletions in a probabilistic modeling framework.** *BMC Bioinformatics* **6**:63.

The Chimpanzee Sequencing and Analysis Consortium. 2005. **Initial sequence of the chimpanzee genome and comparison with the human genome.** *Nature* **437**:69-87.

The International Chimpanzee Chromosome 22 Consotrium. 2004. **DNA sequence and comparative analysis of chimpanzee chromosome 22.** *Nature* **429**:382-388.

Thorne JL, Kishino H, Felsenstein J. 1991. **An evolutionary model for maximum likelihood alignment of DNA sequences.** *J Mol Evol.* **33**:114-124.

Thorne JL, Kishino H, Felsenstein J. 1992. **Inching toward reality: an improved likelihood model of sequence evolution.** *J Mol Evol.* **34**:3-16.

Yamane K, Yano K, Kawahara T. 2006. **Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice.** *DNA Res.* **13**:197-204.

Yang Z. 2006. *Computational Molecular Evolution.* New York (NY), Oxford University Press.

Zhang ZL, Gerstein M. 2003. **Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes.** *Nucleic Acids Res.* **31**:5338-5348.