The Annals of Applied Probability
2015, Vol. 25, No. 3, 1325–1348
DOI: 10.1214/14-AAP1024

THE INTERNAL BRANCH LENGTHS OF THE KINGMAN COALESCENT[1]

BY IULIA DAHMER[2] AND GÖTZ KERSTING

Goethe-Universität Frankfurt

In the Kingman coalescent tree the length of order r is defined as the sum of the lengths of all branches that support r leaves. For $r = 1$ these branches are external, while for $r \geq 2$ they are internal and carry a subtree with r leaves. In this paper we prove that for any $s \in \mathbb{N}$ the vector of rescaled lengths of orders $1 \leq r \leq s$ converges to the multivariate standard normal distribution as the number of leaves of the Kingman coalescent tends to infinity. To this end we use a coupling argument which shows that for any $r \geq 2$ the (internal) length of order r behaves asymptotically in the same way as the length of order 1 (i.e., the external length).

1. Introduction and main result. The Kingman coalescent was introduced in [14] as a model for describing the genealogical relationships between the individuals for a wide class of population models; see [17] for details. The state space of the Kingman n-coalescent, $n \in \mathbb{N}$, is the set $\mathcal{P}_n$ of partitions of the set $\{1, 2, \dots, n\}$. The process starts in the partition into singletons $\pi_n = \{\{1\}, \dots, \{n\}\}$ and has the following dynamics: given that the process is in the state π_k, it jumps after a random time X_k to a state π_{k-1} which is obtained by merging two randomly chosen elements from π_k. The random inter-coalescence times X_k are independent, exponentially distributed random variables with parameters $\binom{k}{2}$. The process can be viewed graphically as a rooted tree that starts from n leaves labelled from 1 to n and whose any two branches coalesce independently at rate 1. Each branch of this tree is situated above a subtree. If this subtree has r leaves, we say that the branch is *of order* r. The branches of order $r \geq 2$ are the *internal* branches, while those of order 1 are the *external* ones (they support subtrees consisting of just one node).

Let us look at the tree from the leaves towards the root (see Figure 2). Then the branch of order r supporting the leaves $i_1, \dots, i_r$ is formed at the level $\sigma(i_1, \dots, i_r)$ and ends at level $\rho(i_1, \dots, i_r)$, where

$$\sigma(i_1, \dots, i_r) = \max\{1 \leq k \leq n : \{i_1, \dots, i_r\} \in \pi_k\}$$

Received April 2013; revised December 2013.

[1]Supported in part by the DFG Priority Programme SPP 1590 “Probabilistic Structures in Evolution.”

[2]Supported by the German Academic Exchange Service (DAAD).

MSC2010 subject classifications. Primary 60K35; secondary 60F05, 60J10.

Key words and phrases. Coalescent, internal branch length, asymptotic distribution, coupling, Markov chain.

and

$$\rho(i_1, \ldots, i_r) = \max\{1 \le k < \sigma(i_1, \ldots, i_r) : \{i_1, \ldots, i_r\} \notin \pi_k\}.$$

For a subset $\{i_1, \ldots, i_r\}$ of leaves, which is not supported by some branch (which means that $\{i_1, \ldots, i_r\} \notin \pi_k$ for all k) we set $\sigma(i_1, \ldots, i_r) = \rho(i_1, \ldots, i_r) = n$.

Let $S_{i_1,\ldots,i_r}$ denote the length of the branch of order r that supports the leaves $i_1, \ldots, i_r$, and write $\mathcal{L}^{n,r}$ for the total length of order r. Then

$$S_{i_1,\ldots,i_r} = \sum_{l=\rho(i_1,\ldots,i_r)+1}^{\sigma(i_1,\ldots,i_r)} X_l$$

and

$$\mathcal{L}^{n,r} = \sum_{1 \le i_1 < \cdots < i_r \le n} S_{i_1,\ldots,i_r}.$$

Observe that $\mathcal{L}^{n,1}$ is the total length of the external branches.

The length of a randomly chosen external branch in the coalescent tree has been studied by Freund and Möhle [8] for the Bolthausen–Sznitman coalescent and by Gnedin et al. [11] for the Λ-coalescent. Asymptotic results concerning the total external length of Beta$(2-\alpha, \alpha)$-coalescents were given by Möhle [16] for the case $0 < \alpha < 1$, by Kersting et al. [13] for the case $\alpha = 1$, and by Kersting et al. [5] for the case $1 < \alpha < 2$. For the case $1 < \alpha < 2$ a weak law of large numbers result concerning $\mathcal{L}^{n,r}$ can be easily deduced from Theorem 9 of Berestycki et al. [2] and also from Dhersin and Yuan [6].

Fu and Li [10] computed the expectation and variance of the total external branch length of the Kingman n-coalescent and Caliebe et al. [4] derived the asymptotic distribution of a randomly chosen external branch. In [12] Janson and Kersting obtained the asymptotic normality of the total external branch length. Our main result states that the same kind of asymptotics holds for the lengths of order $r \ge 1$. Moreover, these lengths are asymptotically independent.

THEOREM. *For any $s \in \mathbb{N}$, as $n \to \infty$*

$$\sqrt{\frac{n}{4\log n}}\left(\mathcal{L}^{n,1} - \mu_1, \ldots, \mathcal{L}^{n,s} - \mu_s\right) \xrightarrow{d} N(0, I_s),$$

where I_s denotes the $s \times s$-identity matrix and $\mu_r = \mathbb{E}(\mathcal{L}^{n,r}) = \frac{2}{r}$ for every $r \ge 1$.

In a forthcoming paper our theorem will be a main building block for proving a functional limit theorem for the total external length of the evolving Kingman-coalescent.

The scatterplot for the lengths of orders 1 and 2 in Figure 1 confirms the theorem. The bulk of the points are located around the mean $(\mu_1, \mu_2) = (2, 1)$. Also,

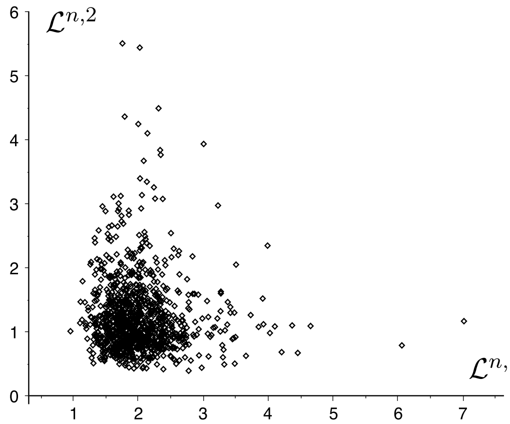

FIG. 1. *External length versus internal length of order* 2. *The plot is based on* 1000 *coalescent realisations with* $n = 100$.

in this region hardly any correlation between the two lengths is visible. The outliers are due to exceptionally long branches whose occurrence has been explained in detail in [12] for the external case. The simulation shows that this phenomenon appears similarly in the case of internal lengths, as one would expect.

As to the proof of the theorem, for the case $s = 1$ a hidden symmetry within the Kingman coalescent is used in [12]. Here we substantially build on the result for $s = 1$; however, the proof for the more general case is rather different. It consists of a coupling device for Markov chains, which connects the total length of order r to the total external length: for $1 \le k \le n$ let $W_k(r)$ denote the *number of order* r *at level* k, the number of branches of order r among the k branches present in the coalescent tree after the $(n-k)$th coalescing event. (Note that here and elsewhere we are suppressing the n in the notation.) That is,

$$W_k(r) := \bigl|\{\{i_1, \dots, i_r\} \subset \{1, \dots, n\} : i_1 < \cdots < i_r, \\ \sigma(i_1, \dots, i_r) \ge k > \rho(i_1, \dots, i_r)\}\bigr|.$$

In particular $W_n(r) = 0, W_{n-1}(r) = 0, \dots, W_{n-r+2}(r) = 0$ and $W_1(r) = 0$ for $r < n$. For an example, see Figure 2.

It is important to notice that for any $s \in \mathbb{N}$, the random vectors $(W_k(1), W_k(2), \dots, W_k(s))$ form a Markov chain if k runs from n to 1 (a property which facilitated our simulations). The transition probabilities of the Markov chain are given explicitly in Section 3. For a similar approach using a Markov chain embedded in the Bolthausen–Sznitman coalescent, see [1]. The idea of our proof is to couple $(W_k(r))_{n \ge k \ge 1}$ for $1 \le r \le s$ jointly with s independent copies of the Markov chain of external numbers $(W_k(1))_{n \ge k \ge 1}$. Since in addition the length of order r is essentially specified by the chain $(W_k(r))_{n \ge k \ge 1}$, it consequently gets the asymptotic behaviour of the external length.

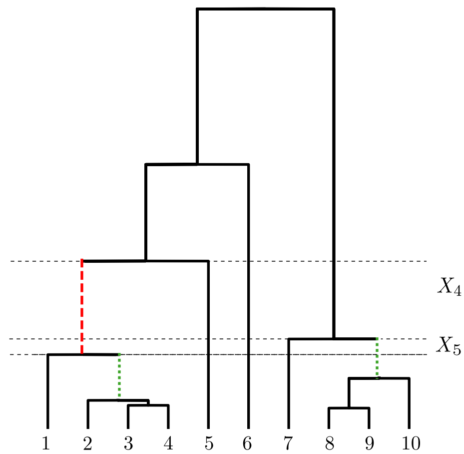

FIG. 2. *The dashed (red) branch is an internal branch of order* 4; *it supports the leaves* 1, 2, 3, *and* 4. *It is formed at level* $\sigma(1,\dots,4) = 5$ *and ends at level* $\rho(1,\dots,4) = 3$. *Its length is* $S_{1,2,3,4} = X_4 + X_5$. *The dotted (green) branches are the branches of order three. The numbers of branches of orders* 1 *to* 10 *at level* 5 *are* $W_5(1) = 3$, $W_5(2) = 0$, $W_5(3) = 1$, $W_5(4) = 1$, *and* $W_5(i) = 0$ *for* $i \geq 5$.

The simulations in Figure 3 give an impression of the behaviour of the lengths of different orders. In the range between the levels n and $n^{1-\varepsilon}$ for small $\varepsilon > 0$ (closer to the leaves) they differ substantially, as seen in Figure 3(a). This deviation is only due to expectations and does not appear at the level of fluctuations. Indeed, it is known from [12] that for the external length the fluctuations are induced just by the $W_k(1)$ with $n^{1-\varepsilon} \geq k \geq \sqrt{n}$. As suggested by Figure 3(b) in this region the

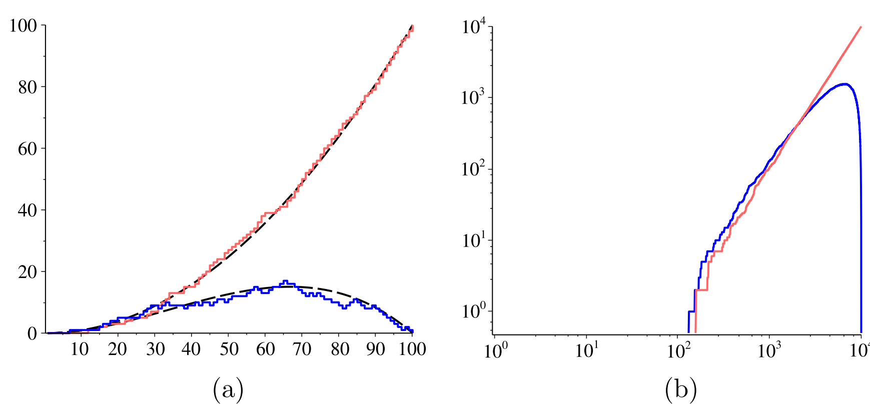

FIG. 3. (a) *Simulations of the external numbers* $W_k(1)$ (*in orange*) *and internal numbers* $W_k(2)$ *of order* 2 (*in blue*) *for a coalescent with* $n = 100$ *for* $1 \leq k \leq n$. *The black dashed curves represent the expectations as given in Lemma* 1. (b) *Gives the representations in double logarithmic scale for a coalescent with* $n = 10^4$.

evolution of the chains is similar for orders $r \geq 2$. The difference in expectation is negligible in our construction, as we couple the jumps of the chains and afterwards consider the lengths of different orders centred at expectation.

The interest in the quantities $\mathcal{L}^{n,r}$ arose from models where the population is subject to mutation, the mutations being modelled as points of a Poisson process with constant rate $\frac{\theta}{2}$ on the branches of the coalescent tree. In the infinitely many sites mutation model, in which every new mutation occurs at a new locus on the DNA, mutations that are located on the external branches of the coalescent tree affect only single individuals, whereas mutations located on an internal branch of order $r \geq 2$ affect all r individuals sitting at the leaves supported by that particular branch. In a population of size n, let $M_r(n)$ denote the number of mutations carried by exactly r individuals. The vector $(M_1(n), \dots, M_{n-1}(n))$, called the site frequency spectrum, and the total number $S_n := \sum_{r=1}^{n-1} M_r(n)$ of mutations that affect the population, called the number of segregating sites, are quantities of statistical importance. Berestycki et al. [2] obtained a weak law of large numbers for $M_r(n)$, $r \geq 1$, in the case of Beta-coalescents with $1 < \alpha < 2$.

For the Kingman coalescent it is known that the number of segregating sites S_n, when rescaled by $\log n$, converges almost surely as $n \to \infty$ to θ; see, for example, [3], Theorem 2.11. The expectation of $M_r(n)$ (which is equal to $\frac{\theta}{r}$), as well as the variances and the covariances of the numbers of mutations $M_r(n)$, were computed by Fu [9] and Durrett [7]. We obtain the following result as a direct consequence of our theorem.

COROLLARY. *For any $s \in \mathbb{N}$, as $n \to \infty$*

$$\big(M_1(n), \dots, M_s(n)\big) \xrightarrow{d} (M_1, \dots, M_s),$$

where $M_1, \dots, M_s$ are independent Poisson-distributed random variables with parameters $\theta\mu_1, \dots, \theta\mu_s$.

For the proof of the corollary, note from the Poissonian structure of the mutation process that the characteristic function of $(M_1(n), \dots, M_s(n))$ is

$$\begin{aligned}\varphi_{(M_1(n),\dots,M_s(n))}(\lambda_1, \dots, \lambda_s) &= \mathbb{E}\big[\mathbb{E}\big[e^{i(\lambda_1 M_1(n) + \cdots + \lambda_s M_s(n))} | \mathcal{T}\big]\big] \\ &= \mathbb{E}\big[e^{\theta \mathcal{L}^{n,1}(e^{i\lambda_1} - 1)} \cdots e^{\theta \mathcal{L}^{n,s}(e^{i\lambda_s} - 1)}\big],\end{aligned}$$

where $\mathcal{T}$ denotes the σ-algebra containing the whole information about the coalescent tree. From our theorem it follows that $\mathcal{L}^{n,r} \xrightarrow{\mathbb{P}} \mu_r$ as $n \to \infty$ and therefore

$$\varphi_{(M_1(n),\dots,M_s(n))}(\lambda_1, \dots, \lambda_s) \longrightarrow e^{\theta\mu_1(e^{i\lambda_1} - 1)} \cdots e^{\theta\mu_s(e^{i\lambda_s} - 1)},$$

as $n \to \infty$.

REMARK. We note that the convergence $\mathcal{L}^{n,r} \xrightarrow{\mathbb{P}} \mu_r$ can also be deduced from the results of Fu [9]: we have that

$$\mathbb{V}(M_r(n)) = \mathbb{V}(\mathbb{E}[M_r(n)|\mathcal{T}]) + \mathbb{E}(\mathbb{V}[M_r(n)|\mathcal{T}]) = \mathbb{V}\Big(\frac{\theta}{2}\mathcal{L}^{n,r}\Big) + \mathbb{E}\Big(\frac{\theta}{2}\mathcal{L}^{n,r}\Big).$$

Comparing this with Fu's formulas (1)–(3), we obtain for $r < \frac{n}{2}$ that

$$\mathbb{V}(\mathcal{L}^{n,r}) = \frac{2n}{(n-r)(n-r-1)} \sum_{i=r+1}^{n} \frac{1}{i} - \frac{2}{n-r-1}.$$

In particular $\mathbb{V}(\mathcal{L}^{n,r}) \to 0$ and $\mathcal{L}^{n,r} \xrightarrow{\mathbb{P}} \mathbb{E}(\mathcal{L}^{n,r}) = \mu_r$ as $n \to \infty$.

NOTATION. We use the notation $X_n = O_P(f(n))$ for $f(n) > 0$ if

$$\lim_{a\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(X_n > a \cdot f(n)\big) = 0,$$

that is, $\frac{X_n}{f(n)}$ is stochastically bounded.

Throughout c denotes a finite constant whose value is not important and may change from line to line.

2. Moment computations.

LEMMA 1. *For the expectation and variance of $W_k(r)$ the following is true. For $n > r$,*

$$\mathbb{E}\big(W_k(r)\big) = \frac{(n-k)\cdots(n-k-r+2)}{(n-1)\cdots(n-r)} \cdot k(k-1) \quad \textit{and} \quad \mathbb{V}\big(W_k(r)\big) \le c\frac{k^2}{n},$$

where $c < \infty$ is a constant depending on r. In particular

$$\mathbb{E}\big(W_k(r)\big) = \Big(\frac{n-k}{n}\Big)^{r-1} \cdot \frac{k^2}{n} + O\Big(\frac{k}{n}\Big) = \frac{k^2}{n} + O\Big(\frac{k^3}{n^2} + \frac{k}{n}\Big) = O\Big(\frac{k^2}{n}\Big)$$

uniformly in $k \le n$. Also, for any integer $\alpha \ge 2$,

$$\mathbb{E}\big(W_k^\alpha(r)\big) = \Big(\Big(\frac{n-k}{n}\Big)^{r-1} \cdot \frac{k^2}{n}\Big)^\alpha + O\Big(\Big(\frac{k^2}{n}\Big)^{\alpha-1} + \frac{k^2}{n}\Big) = O\Big(\frac{k^{2\alpha}}{n^\alpha} + \frac{k^2}{n}\Big).$$

PROOF. In order to compute the moments of $W_k(r)$, let us again label the leaves of the coalescent tree from 1 to n and note that $W_k(r)$ can be written as

$$W_k(r) = \sum_{1\le i_1<\cdots<i_r\le n} \mathbf{1}_{\{\{i_1,\ldots,i_r\}\in\pi_k\}},$$

where π_k is the state of the coalescent process at time k. Then for $n > r$, using the fact that the event $\{\{1,\ldots,r\} \in \pi_k\}$ is the disjoint union (over $n > l_1 > l_2 > \cdots >$

$l_{r-1} \geq k$) of the events {the branch supporting leaves $1, \dots, r$ is formed by $r-1$ coalescing events happening at levels $l_1, l_2, \dots, l_{r-1}$}, we have from exchangeability that

$$
\begin{aligned}
\mathbb{E}(W_k(r)) &= \mathbb{E}\biggl(\sum_{1 \leq i_1 < \cdots < i_r \leq n} \mathbf{1}_{\{\{i_1,\dots,i_r\}\in\pi_k\}}\biggr) = \binom{n}{r} \mathbb{P}(\{1,\dots,r\} \in \pi_k) \\
&= \binom{n}{r} \sum_{n > l_1 > l_2 > \cdots > l_{r-1} \geq k} \frac{\binom{n-r}{2}}{\binom{n}{2}} \cdot \frac{\binom{n-1-r}{2}}{\binom{n-1}{2}} \cdots \\
&\qquad \times \frac{\binom{l_1+2-r}{2}}{\binom{l_1+2}{2}} \cdot \frac{\binom{r}{2}}{\binom{l_1+1}{2}} \cdot \frac{\binom{l_1-(r-1)}{2}}{\binom{l_1}{2}} \cdots \qquad (1) \\
&\qquad \times \frac{\binom{l_j+2-(r-j+1)}{2}}{\binom{l_j+2}{2}} \cdot \frac{\binom{r-j+1}{2}}{\binom{l_j+1}{2}} \cdot \frac{\binom{l_j-(r-j)}{2}}{\binom{l_j}{2}} \cdots \\
&\qquad \times \frac{\binom{l_{r-1}}{2}}{\binom{l_{r-1}+2}{2}} \cdot \frac{\binom{2}{2}}{\binom{l_{r-1}+1}{2}} \cdot \frac{\binom{l_{r-1}-1}{2}}{\binom{l_{r-1}}{2}} \cdots \frac{\binom{k}{2}}{\binom{k+1}{2}}.
\end{aligned}
$$

Most binomials in the nominator and the denominator cancel. The summands turn out to be equal such that

$$
\begin{aligned}
\mathbb{E}(W_k(r)) &= \binom{n}{r} \sum_{n > l_1 > l_2 > \cdots > l_{r-1} \geq k} \frac{\binom{r}{2} \cdots \binom{2}{2}}{\binom{n}{2} \cdots \binom{n-r+1}{2}} \cdot \binom{k}{2} \\
&= \binom{n}{r} \binom{n-k}{r-1} \frac{\binom{r}{2} \cdots \binom{2}{2}}{\binom{n}{2} \cdots \binom{n-r+1}{2}} \cdot \binom{k}{2} \\
&= \frac{(n-k) \cdots (n-k-r+2)}{(n-1) \cdots (n-r)} \cdot k(k-1).
\end{aligned}
$$

This is the first claim, which directly implies the first asymptotic formula for $\mathbb{E}(W_k(r))$. Now the second follows by means of the Bernoulli inequality:

$$
1 - \left(\frac{n-k}{n}\right)^{r-1} = 1 - \left(1 - \frac{k}{n}\right)^{r-1} \leq (r-1)\frac{k}{n}.
$$

The computation of the second moment of $W_k(r)$ follows in a similar way. Note that the event $\{\{i_1, \dots, i_r\}, \{j_1, \dots, j_r\} \in \pi_k\}$ is nonempty only if the sets $\{i_1, \dots, i_r\}$ and $\{j_1, \dots, j_r\}$ are identical or disjoint. Thus, for $n > 2r$

$$
\begin{aligned}
\mathbb{E}(W_k^2(r)) &= \binom{n}{r} \mathbb{E}(\mathbf{1}^2_{\{\{1,\dots,r\}\in\pi_k\}}) \\
&\qquad + \binom{n}{r, r, n-2r} \mathbb{E}(\mathbf{1}_{\{\{1,\dots,r\}\in\pi_k\}} \cdot \mathbf{1}_{\{\{r+1,\dots,2r\}\in\pi_k\}})
\end{aligned}
$$

$$
\begin{aligned}
&= \binom{n}{r} \mathbb{P}(\{1,\dots,r\} \in \pi_k) \\
&\quad + \binom{n}{r,r,n-2r} \mathbb{P}(\{1,\dots,r\},\{r+1,\dots,2r\} \in \pi_k) \\
&= \mathbb{E}(W_k(r)) \\
&\quad + \binom{n}{r,r,n-2r} \sum \frac{\binom{n-2r}{2}}{\binom{n}{2}} \cdots \frac{\binom{l_1''+2-2r}{2}}{\binom{l_1''+2}{2}} \cdot \frac{\binom{r}{2}}{\binom{l_1''+1}{2}} \cdot \frac{\binom{l_1''-(2r-1)}{2}}{\binom{l_1''}{2}} \cdots \\
&\qquad \times \frac{\binom{l_{2r-2}''-1}{2}}{\binom{l_{2r-2}''+2}{2}} \cdot \frac{1}{\binom{l_{2r-2}''+1}{2}} \cdot \frac{\binom{l_{2r-2}''-2}{2}}{\binom{l_{2r-2}''-2}{2}} \cdots \frac{\binom{k-1}{2}}{\binom{k+1}{2}},
\end{aligned}
$$

where the sum is taken over all $n > l_1 > l_2 > \cdots > l_{r-1} \geq k$ and all $n > l_1' > l_2' > \cdots > l_{r-1}' \geq k$ such that $\{l_1,\dots,l_{r-1}\} \cap \{l_1',\dots,l_{r-1}'\} = \varnothing$. The sequences $(l_j)_{1\leq j\leq r-1}$ and $(l_j')_{1\leq j\leq r-1}$ denote the coalescence times of the branches supporting leaves from the sets $\{1,\dots,r\}$ and $\{r+1,\dots,2r\}$, respectively. The sequence $(l_j'')_{1\leq j\leq 2r-2}$ is the reordering of $l_1,\dots,l_{r-1},l_1',\dots,l_{r-1}'$ in decreasing order. Thus

$$
\begin{aligned}
\mathbb{E}(W_k^2(r)) &= \mathbb{E}(W_k(r)) \\
&\quad + \binom{n}{r,r,n-2r} \binom{n-k}{r-1,r-1,n-k-2r+2} \frac{(\binom{r}{2}\cdots\binom{2}{2})^2}{\binom{n}{2}\cdots\binom{n-2r+1}{2}} \\
&\quad \times \binom{k}{2}\binom{k-1}{2} \\
&= \mathbb{E}(W_k(r)) \\
&\quad + \frac{(n-k)\cdots(n-k-2r+3)}{(n-1)\cdots(n-2r)} k(k-1)^2(k-2).
\end{aligned}
$$

The variance of W_k is then for $k \leq n-1$

$$
\begin{aligned}
\mathbb{V}(W_k(r)) &= \mathbb{E}(W_k(r)) \biggl(1 + \frac{(n-k-r+1)\cdots(n-k-2r+3)}{(n-r-1)\cdots(n-2r)}(k-1)(k-2) \\
&\qquad - \frac{(n-k)\cdots(n-k-r+2)}{(n-1)\cdots(n-r)} \cdot k(k-1)\biggr) \\
&\leq \mathbb{E}(W_k(r)) \biggl(1 + k(k-1)(n-k)\cdots(n-k-r+2) \\
&\qquad \times \biggl(\frac{1}{(n-r-1)\cdots(n-2r)} - \frac{1}{(n-1)\cdots(n-r)}\biggr)\biggr) \\
&\leq \mathbb{E}(W_k(r)) \biggl(1 + k^2 n^{r-1}\biggl(\frac{1}{(n-2r)^r} - \frac{1}{n^r}\biggr)\biggr).
\end{aligned}
$$

Using the mean value theorem we obtain that

$$\mathbb{V}\big(W_k(r)\big) \le \mathbb{E}\big(W_k(r)\big)\bigg(1 + k^2 n^{r-1} \frac{2r^2}{(n-2r)^{r+1}}\bigg) \le c\frac{k^2}{n},$$

for $c < \infty$ depending on r.

For the other claims we use the same type of argument as above. We have that

$$\begin{aligned}
&\binom{n}{r,\dots,r,n-\alpha r}\mathbb{P}\big(\{1,\dots,r\},\dots,\{(\alpha-1)r+1,\dots,\alpha r\} \in \pi_k\big)\\
&\quad= \binom{n}{r,\dots,r,n-\alpha r}\binom{n-k}{r-1,\dots,r-1,n-k-\alpha(r-1)}\\
&\qquad\times \frac{(\binom{r}{2}\cdots\binom{2}{2})^\alpha}{\binom{n}{2}\cdots\binom{n-\alpha r+1}{2}}\cdot\binom{k}{2}\cdots\binom{k-\alpha+1}{2}\\
&\quad= \frac{(n-k)\cdots(n-k-\alpha(r-1)+1)}{(n-1)\cdots(n-\alpha r)}\\
&\qquad\times k(k-1)^2\cdots(k-\alpha+1)^2(k-\alpha)\\
&\quad= \frac{(n-k)^{\alpha(r-1)}}{n^{\alpha r}}\cdot k^{2\alpha} + O\bigg(\frac{(n-k)^{\alpha(r-1)}}{n^{\alpha r}}\cdot k^{2\alpha-1}\bigg)\\
&\qquad+ O\bigg(\frac{(n-k)^{\alpha(r-1)-1}}{n^{\alpha r}}\cdot k^{2\alpha}\bigg) + O\bigg(\frac{(n-k)^{\alpha(r-1)}}{n^{\alpha r+1}}\cdot k^{2\alpha}\bigg)\\
&\quad= \frac{(n-k)^{\alpha(r-1)}}{n^{\alpha r}}\cdot k^{2\alpha} + O\bigg(\frac{k^{2\alpha-1}}{n^\alpha}\bigg).
\end{aligned} \tag{2}$$

In particular this gives the asymptotic expansion of $\mathbb{E}(W_k(r))$. Also

$$\begin{aligned}
&\binom{n}{r,\dots,r,n-\alpha r}\mathbb{P}\big(\{1,\dots,r\},\dots,\{(\alpha-1)r+1,\dots,\alpha r\} \in \pi_k\big)\\
&\quad= O\bigg(\frac{k^{2\alpha}}{n^\alpha}\bigg).
\end{aligned} \tag{3}$$

Moreover, by expanding $(\sum_{1\le i_1<\cdots<i_r\le n} \mathbf{1}_{\{\{i_1,\dots,i_r\}\in\pi_k\}})^\alpha$

$$\begin{aligned}
\mathbb{E}\big(W_k^\alpha(r)\big) &= \binom{n}{r,\dots,r,n-\alpha r}\mathbb{P}\big(\{1,\dots,r\},\dots,\{(\alpha-1)r+1,\dots,\alpha r\} \in \pi_k\big)\\
&\quad+ O\Bigg(\sum_{\beta=1}^{\alpha-1}\binom{n}{r,\dots,r,n-\beta r}\\
&\qquad\times \mathbb{P}\big(\{1,\dots,r\},\dots,\{(\beta-1)r+1,\dots,\beta r\} \in \pi_k\big)\Bigg).
\end{aligned} \tag{4}$$

The last claim now follows from (2), (3), (4), and the fact that $\sum_{\beta=1}^{\alpha-1}(\frac{k^2}{n})^\beta = O((\frac{k^2}{n})^{\alpha-1}+\frac{k^2}{n})$. $\square$

REMARK. It can be read off from the computation in (1) that in the case $r=2$, given the event $\{\rho(1,2)<k\}=\{\{1,2\}\in\pi_k\}$, the random variable $\sigma(1,2)$ is uniformly distributed on the set of levels $\{k,\dots,n-1\}$. Indeed for $k\le l<n$ it holds

$$\begin{aligned}&\mathbb{P}\big(\sigma(1,2)=l,\rho(1,2)<k\big)\\&\quad=\frac{\binom{n-2}{2}}{\binom{n}{2}}\cdot\frac{\binom{n-3}{2}}{\binom{n-1}{2}}\cdots\frac{\binom{l+1}{2}}{\binom{l+3}{2}}\cdot\frac{\binom{l}{2}}{\binom{l+2}{2}}\cdot\frac{1}{\binom{l+1}{2}}\cdot\frac{\binom{l-1}{2}}{\binom{l}{2}}\cdot\frac{\binom{l-2}{2}}{\binom{l-1}{2}}\cdots\frac{\binom{k}{2}}{\binom{k+1}{2}}\\&\quad=\frac{1}{\binom{n}{2}}\cdot\frac{1}{\binom{n-1}{2}}\cdot\binom{k}{2},\end{aligned}$$

which does not depend on l. A similar observation can be made for the case $r>2$.

Using the numbers $W_k(r)$ we have the following simplified expression for the length of order r:

$$\mathcal{L}^{n,r}=\sum_{2\le k\le n}W_k(r)\cdot X_k. \tag{5}$$

We note from this representation that there are two sources of randomness in the length of order r, one coming from the numbers $W_k(r)$ and one coming from the exponential inter-coalescence times. It is easy to see that taking out the randomness introduced by the exponential times (i.e., replacing them by their expectations) leads to an error that is asymptotically $O_P(n^{-1/2})$ and therefore converges to 0 after the rescaling by $\sqrt{\frac{n}{\log n}}$. Indeed, by using the independence between the X_k's and the $W_k(r)$'s and Lemma 1, we have that for a constant $c<\infty$

$$\begin{aligned}&\mathbb{V}\Big(\sum_{2\le k\le n}W_k(r)\cdot\big(X_k-\mathbb{E}(X_k)\big)\Big)\\&\quad=\sum_{2\le k\le n}\mathbb{V}\big(W_k(r)\cdot\big(X_k-\mathbb{E}(X_k)\big)\big)\\&\quad=\sum_{2\le k\le n}\mathbb{E}\big(W_k^2(r)\big)\cdot\mathbb{E}\big(\big(X_k-\mathbb{E}(X_k)\big)^2\big)\\&\quad\le c\sum_{2\le k\le n}\Big(\frac{k^4}{n^2}+\frac{k^2}{n}\Big)\cdot\frac{1}{\binom{k}{2}^2}\le c\frac{1}{n}\end{aligned}$$

and therefore

$$\mathcal{L}^{n,r}=L^{n,r}+O_P\big(n^{-1/2}\big),$$

where

$$L^{n,r} := \sum_{2\le k\le n} W_k(r)\cdot \mathbb{E}(X_k) = \sum_{2\le k\le n} W_k(r)\cdot \frac{2}{k(k-1)}. \tag{6}$$

As a consequence, in the proof of our theorem we need only focus on the length $L^{n,r}$ which we will, for convenience, still call the length of order r.

3. The coupling. Our proof follows a coupling argument which substantially relies on the observation that for every $s \in \mathbb{N}$ the vector $V_k := (W_k(1), W_k(2), \dots, W_k(s))$ follows for $n \ge k \ge 1$ the dynamics of an inhomogenous Markov chain with state space $\mathcal{X}_{n,s} := \{0, 1, \dots, n\}^s$. We let time run in coalescent direction (from the leaves to the root of the tree), and for convenience we consider the evolution of the chain $(V_k)_{n\ge k\ge 1}$, running in the same direction, namely from level n to level 1.

For every $1 \le r \le s$ we denote by $\Delta W_{n-1}(r), \dots, \Delta W_1(r)$ the sizes of the jumps of the chain $(W_k(r))_{n\ge k\ge 1}$,

$$\Delta W_k(r) := W_k(r) - W_{k+1}(r), \qquad n-1 \ge k \ge 1$$

and observe that $\Delta W_k(r) \in \{-2, -1, 0, 1\}$ for all k. The jumps of size 1 correspond to the levels at which a new branch of order r is formed (by the coalescence of two other branches), whereas the jumps of sizes -1 and -2 happen at the levels at which one (or, resp., two) branches of order r end (by coalescence of one of them with some other branch or by mutual coalescence).

For $1 \le k \le n$ and $v, v' \in \mathcal{X}_{n,s}$ let

$$P_v^k(v') := \mathbb{P}(V_{k-1} = v' | V_k = v)$$

denote the transition probabilities of $(V_k)_{n\ge k\ge 1}$. They are given for $v = (w_1, \dots, w_s)$, $w_1 + \cdots + w_s \le k$ by

$$P_v^k(v') = \begin{cases} \dfrac{\binom{k-w_1-\cdots-w_s}{2}}{\binom{k}{2}}, & \\ & \text{if } v = v', \\ \dfrac{w_i(k-\sum_{i=1}^s w_i)}{\binom{k}{2}}, & \\ & \text{if } v = v' - e_i \text{ for some } i, \\ \dfrac{\binom{w_i}{2}}{\binom{k}{2}}, & \text{if } v = v' - 2e_i + e_{2i} \text{ for some } i, \\ \dfrac{w_i w_j}{\binom{k}{2}}, & \text{if } v = v' - e_i - e_j + e_{i+j} \text{ for some } i \ne j, \\ 0, & \text{else,} \end{cases} \tag{7}$$

where $e_i = (\delta_{i,l})_{1\le l\le s}$ [note that $e_i = (0,\dots,0)$ for $i > s$]. For all other $v \in \mathcal{X}_{n,s}$ we set for definiteness $P_v^k(v') = \delta_{v,v'}$ in order to obtain a proper transition matrix.

In particular for $s = 2$ the transition probabilities in (7) are given for $w_1 + w_2 \le k$ by

$$\begin{aligned}
P^k_{(w_1,w_2)}(w_1, w_2) &= \frac{\binom{k-w_1-w_2}{2}}{\binom{k}{2}},\\
P^k_{(w_1,w_2)}(w_1-1, w_2-1) &= \frac{w_1 w_2}{\binom{k}{2}},\\
P^k_{(w_1,w_2)}(w_1-1, w_2) &= \frac{w_1(k-w_1-w_2)}{\binom{k}{2}},\\
P^k_{(w_1,w_2)}(w_1-2, w_2+1) &= \frac{\binom{w_1}{2}}{\binom{k}{2}},\\
P^k_{(w_1,w_2)}(w_1, w_2-1) &= \frac{w_2(k-w_1-w_2)}{\binom{k}{2}},\\
P^k_{(w_1,w_2)}(w_1, w_2-2) &= \frac{\binom{w_2}{2}}{\binom{k}{2}}
\end{aligned}$$

and for $s = 1$ for $w \le k$ by

$$\begin{aligned}
&P^k_w(w) = \frac{\binom{k-w}{2}}{\binom{k}{2}}, \qquad P^k_w(w-1) = \frac{w(k-w)}{\binom{k}{2}},\\
&P^k_w(w-2) = \frac{\binom{w}{2}}{\binom{k}{2}}.
\end{aligned} \tag{8}$$

Let us now describe the coupling in detail. Let $1 < a_n \le n$ and $s \in \mathbb{N}$ be fixed. Starting at a_n we couple the Markov chain $(V_k)_{a_n\ge k\ge 1}$, $V_k = (W_k(1),\dots,W_k(s))$, with another chain $(\widetilde{V}_k)_{a_n\ge k\ge 1}$, $\widetilde{V}_k = (\widetilde{W}_k(1),\dots,\widetilde{W}_k(s))$, defined on the same probability space as $(V_k)_{a_n\ge k\ge 1}$ and having the same state space $\mathcal{X}_{n,s}$. The components of $(\widetilde{V}_k)_{a_n\ge k\ge 1}$ evolve as independent copies of $(W_k(1))_{a_n\ge k\ge 1}$, the Markov chain of external numbers. Therefore its transition probabilities $\widetilde{P}^k_v(\cdot)$ for $v \in \mathcal{X}_{n,s}$ and $1 < k \le a_n$ are given by the product of the transition probabilities of its s components, given in (8). The process $((V_k, \widetilde{V}_k))_{a_n\ge k\ge 1}$ is constructed as a Markov chain, where the jumps are coupled in a way that we will describe in detail shortly.

Thus let Q^k_v and $\widetilde{Q}^k_{\widetilde{v}}$ denote the conditional distributions of the jumps ΔV_k and $\Delta\widetilde{V}_k$ of the two Markov chains, given the current states v and $\widetilde{v}$, respectively. (The notation ΔV_k and $\Delta\widetilde{V}_k$ refer to component-wise differences.) For the sequel it is important that the leading terms of Q^k_v and $\widetilde{Q}^k_{\widetilde{v}}$ agree. More precisely, from (7) and (8), under the constrain that $w_1 + \cdots + w_s \le k$

$$
(9) \quad Q_v^k(z), \widetilde{Q}_v^k(z) = \begin{cases} 1 - \sum_{j=1}^{s} \frac{2w_j}{k} + O\left(\sum_{j=1}^{s} \frac{w_j^2}{k^2}\right), & \text{if } z = (0, \dots, 0), \\ \frac{2w_i}{k} + O\left(\sum_{j=1}^{s} \frac{w_j^2}{k^2}\right), & \text{if } z = -e_i \text{ for some } i, \\ O\left(\sum_{j=1}^{s} \frac{w_j^2}{k^2}\right), & \text{else,} \end{cases}
$$

where $z \in \{-2, -1, 0, 1\}^s$ and $e_i = (\delta_{i,l})_{1 \le l \le s}$. Here we use that $w_i w_j \le w_i^2 + w_j^2$ and $w_i \le k$.

As it is well known (see, e.g., [15]), an optimal coupling of the two distributions is specified as follows. Let $\|\cdot\|_{\mathrm{TV}}$ denote the total variation distance between two distributions, and define

$$
p = p_{v,\tilde{v}} := 1 - \|Q_v^k - \widetilde{Q}_{\tilde{v}}^k\|_{\mathrm{TV}}.
$$

Then, with probability p choose $\Delta V_k = \Delta \widetilde{V}_k = Z$, where the random variable Z has distribution γ_I, given by its weights

$$
\gamma_I(z) = \frac{Q_v^k(z) \wedge \widetilde{Q}_{\tilde{v}}^k(z)}{p},
$$

$z \in \{-2, -1, 0, 1\}^s$, and with probability $1 - p$ choose ΔV_k according to the probability distribution weights

$$
\gamma_{II}(z) = \frac{(Q_v^k(z) - \widetilde{Q}_{\tilde{v}}(z))^+}{1 - p},
$$

and independently choose $\Delta \widetilde{V}_k$ according to the probability distribution weights

$$
\gamma_{III}(z) = \frac{(\widetilde{Q}_{\tilde{v}}(z) - Q_v^k(z))^+}{1 - p},
$$

$z \in \{-2, -1, 0, 1\}^s$.

This coupling is optimal in the sense that the probability $\mathbb{P}(\Delta V_k \neq \Delta \widetilde{V}_k | V_k = v, \widetilde{V}_k = \tilde{v})$ is minimal among the corresponding probabilities for couplings of the two distributions Q_v^k and $\widetilde{Q}_{\tilde{v}}^k$, and therefore it is equal to $\|Q_v^k - \widetilde{Q}_{\tilde{v}}^k\|_{\mathrm{TV}}$. As starting distribution of the coupled chain $(V_k, \widetilde{V}_k)_{a_n \ge k \ge 1}$ we allow any distribution of $(V_{a_n}, \widetilde{V}_{a_n})$ such that the marginals are the distributions of V_{a_n} and $\widetilde{V}_{a_n}$, respectively. We point out that the distributions of V_{a_n} and $\widetilde{V}_{a_n}$ are given by the Kingman coalescent at level a_n. Up to this constraint the common distribution is arbitrary.

The next two lemmas give essential properties of the coupling.

LEMMA 2. *There is a $c < \infty$ such that the above defined coupling satisfies for $r \le s$*

$$\mathbb{P}\big(\Delta W_k(r) \ne \Delta \widetilde{W}_k(r)\big) \le c\left(\frac{k}{a_n\sqrt{n}} + \frac{a_n k}{n^2} + \frac{1}{k}\right)$$

and

$$\mathbb{E}\big(\big|W_k(r) - \widetilde{W}_k(r)\big|\big) \le c\left(\frac{k^2}{a_n\sqrt{n}} + \frac{a_n k^2}{n^2} + 1\right)$$

for all $1 \le k < a_n$.

PROOF. In the proof we write as an abbreviation W_k instead of $W_k(r)$ and similarly $\widetilde{W}_k$, ΔW_k and $\Delta\widetilde{W}_k$ instead of $\widetilde{W}_k(r)$, $\Delta W_k(r)$ and $\Delta\widetilde{W}_k(r)$, respectively.

From (9) it follows that for both the chains $(V_k)_{a_n \ge k \ge 1}$ and $(\widetilde{V}_k)_{a_n \ge k \ge 1}$ jumps of sizes $(0, \dots, 0)$ and $-e_i$ with $1 \le i \le r$ occur with probabilities of larger order than jumps of other sizes. It follows from the coupling that

$$\begin{aligned}\{\Delta W_k \ne \Delta\widetilde{W}_k\} \subset \{&\Delta W_k = -1, \Delta V_k \ne \Delta\widetilde{V}_k\}\\ &\cup \{\Delta\widetilde{W}_k = -1, \Delta V_k \ne \Delta\widetilde{V}_k\}\\ &\cup \big\{\Delta W_k \in \{1, -2\}\big\} \cup \{\Delta\widetilde{W}_k = -2\}.\end{aligned}$$

Note that since $\widetilde{W}_k$ has the distribution of the external number at level k, the jumps size $\Delta\widetilde{W}_k$ cannot take the value 1.

Thus, writing as an abbreviation $\mathbb{P}^k_{v,\widetilde{v}}(\cdot)$ for the conditional probability given the event $\{V_k = v, \widetilde{V}_k = \widetilde{v}\}$, we obtain

$$\begin{aligned}\mathbb{P}^{k+1}_{v,\widetilde{v}}(\Delta W_k \ne \Delta\widetilde{W}_k) &\le (1-p)\gamma_{II}(\Delta W_k = -1) + (1-p)\gamma_{III}(\Delta\widetilde{W}_k = -1)\\ &\quad + \mathbb{P}^{k+1}_{v,\widetilde{v}}\big(\Delta W_k \in \{1,-2\}\big) + \mathbb{P}^{k+1}_{v,\widetilde{v}}(\Delta\widetilde{W}_k = -2)\\ &\le (1-p)\gamma_{II}(\Delta V_k = -e_r) + (1-p)\gamma_{III}(\Delta\widetilde{V}_k = -e_r)\\ &\quad + c\cdot\sum_{i=1}^{s}\frac{w_i^2 + \widetilde{w}_i^2}{k^2}\\ &\quad + \mathbb{P}^{k+1}_{v,\widetilde{v}}\big(\Delta W_k \in \{1,-2\}\big) + \mathbb{P}^{k+1}_{v,\widetilde{v}}(\Delta\widetilde{W}_k = -2)\\ &\le (1-p)\gamma_{II}(\Delta V_k = -e_r) + (1-p)\gamma_{III}(\Delta\widetilde{V}_k = -e_r)\\ &\quad + c\cdot\sum_{i=1}^{s}\frac{w_i^2 + \widetilde{w}_i^2}{k^2}.\end{aligned}$$

Using now the definitions of γ_{II} and γ_{III} we get that

$$\mathbb{P}^{k+1}_{v,\widetilde{v}}(\Delta W_k \ne \Delta\widetilde{W}_k) \le \big|Q^{k+1}_v(-e_r) - \widetilde{Q}^{k+1}_{\widetilde{v}}(-e_r)\big| + c\cdot\sum_{i=1}^{s}\frac{w_i^2 + \widetilde{w}_i^2}{k^2}. \tag{10}$$

Let us introduce the filtration $\mathbb{F} = (\mathcal{F}_k)_{1\le k\le a_n}$ with $\mathcal{F}_{a_n} \subset \mathcal{F}_{a_n-1} \subset \cdots \subset \mathcal{F}_1$ defined by

$$\mathcal{F}_k = \sigma\big((V_j)_{k\le j\le a_n}, (\widetilde{V}_j)_{k\le j\le a_n}\big).$$

Then (10) in view of (9) may be written as

$$\begin{aligned}\mathbb{P}(\Delta W_k \neq \Delta \widetilde{W}_k|\mathcal{F}_{k+1}) \\ \le \frac{2}{k}|W_{k+1} - \widetilde{W}_{k+1}| + c\sum_{i=1}^{s}\frac{W_{k+1}^2(i) + \widetilde{W}_{k+1}^2(i)}{k^2}\end{aligned} \tag{11}$$

for a constant $c < \infty$. Taking expectation in the inequality above, we obtain using Lemma 1

$$\mathbb{P}(\Delta W_k \neq \Delta \widetilde{W}_k) \le \frac{2}{k}\cdot \mathbb{E}\big(|W_{k+1} - \widetilde{W}_{k+1}|\big) + c\bigg(\frac{k^2}{n^2} + \frac{1}{n}\bigg). \tag{12}$$

We now proceed to finding a bound for $\mathbb{E}(|W_k - \widetilde{W}_k|)$ for $2 \le k \le a_n$. From the transition probabilities (7) we get that

$$\begin{aligned}\mathbb{E}[\Delta W_k|\mathcal{F}_{k+1}] &= (-1)\cdot\frac{W_{k+1}(r)(k+1-W_{k+1}(1)-\cdots-W_{k+1}(r))}{\binom{k+1}{2}} \\ &\quad + (-1)\cdot\frac{W_{k+1}(r)W_{k+1}(1)+\cdots+W_{k+1}(r)W_{k+1}(r-1)}{\binom{k+1}{2}} \\ &\quad + (-2)\cdot\frac{\binom{W_{k+1}(r)}{2}}{\binom{k+1}{2}} + 1\cdot\frac{Z_{k+1}}{\binom{k+1}{2}} \\ &= -\frac{2}{k+1}W_{k+1} + \frac{Z_{k+1}}{\binom{k+1}{2}},\end{aligned}$$

where, letting $d_r = 1$ if r is even and 0 otherwise,

$$Z_k = Z_k(r) := \sum_{\substack{1\le i\le r-1\\ i\neq r-i}} W_k(i)W_k(r-i) + d_r\cdot\binom{W_k(r/2)}{2}. \tag{13}$$

Therefore

$$\mathbb{E}[\Delta W_k|\mathcal{F}_{k+1}] = -\frac{2}{k+1}W_{k+1} + \frac{Z_{k+1}}{\binom{k+1}{2}} \tag{14}$$

and also with a similar but even simpler calculation using (8),

$$\mathbb{E}[\Delta \widetilde{W}_k|\mathcal{F}_{k+1}] = -\frac{2}{k+1}\widetilde{W}_{k+1}. \tag{15}$$

Now note that the absolute value of the difference between the jumps of W_{k+1} and $\widetilde{W}_{k+1}$ is at most 3. Thus

$$\begin{aligned}
&\mathbb{E}\big[|W_k - \widetilde{W}_k||\mathcal{F}_{k+1}\big] \\
&\quad = \mathbb{E}\big[|W_{k+1} - \widetilde{W}_{k+1} + \Delta W_k - \Delta \widetilde{W}_k||\mathcal{F}_{k+1}\big] \\
&\quad \le \mathbb{E}[W_{k+1} - \widetilde{W}_{k+1} + \Delta W_k - \Delta \widetilde{W}_k|\mathcal{F}_{k+1}] \cdot \mathbf{1}_{\{W_{k+1} - \widetilde{W}_{k+1} \ge 3\}} \\
&\qquad + \mathbb{E}[\widetilde{W}_{k+1} - W_{k+1} + \Delta \widetilde{W}_k - \Delta W_k|\mathcal{F}_{k+1}] \cdot \mathbf{1}_{\{W_{k+1} - \widetilde{W}_{k+1} \le -3\}} \\
&\qquad + \big(|W_{k+1} - \widetilde{W}_{k+1}| + \mathbb{E}\big[|\Delta W_k - \Delta \widetilde{W}_k||\mathcal{F}_{k+1}\big]\big) \cdot \mathbf{1}_{\{|W_{k+1} - \widetilde{W}_{k+1}| \le 2\}}.
\end{aligned}$$

Using (14) and (15) we obtain

$$\begin{aligned}
&\mathbb{E}\big[|W_k - \widetilde{W}_k||\mathcal{F}_{k+1}\big] \\
&\quad \le \left(W_{k+1} - \widetilde{W}_{k+1} - \frac{2}{k+1}(W_{k+1} - \widetilde{W}_{k+1}) + \frac{Z_{k+1}}{\binom{k+1}{2}}\right) \cdot \mathbf{1}_{\{W_{k+1} - \widetilde{W}_{k+1} \ge 3\}} \\
&\qquad + \left(\widetilde{W}_{k+1} - W_{k+1} - \frac{2}{k+1}(\widetilde{W}_{k+1} - W_{k+1}) - \frac{Z_{k+1}}{\binom{k+1}{2}}\right) \cdot \mathbf{1}_{\{W_{k+1} - \widetilde{W}_{k+1} \le -3\}} \\
&\qquad + \big(|W_{k+1} - \widetilde{W}_{k+1}| + 3 \cdot \mathbb{P}(\Delta W_k \ne \Delta \widetilde{W}_k|\mathcal{F}_{k+1})\big) \cdot \mathbf{1}_{\{|W_{k+1} - \widetilde{W}_{k+1}| \le 2\}}.
\end{aligned}$$

By (11) we have that

$$\begin{aligned}
&\mathbb{E}\big[|W_k - \widetilde{W}_k||\mathcal{F}_{k+1}\big] \\
&\quad \le \left(|W_{k+1} - \widetilde{W}_{k+1}| - \frac{2}{k+1}|W_{k+1} - \widetilde{W}_{k+1}| + \frac{Z_{k+1}}{\binom{k+1}{2}}\right) \cdot \mathbf{1}_{\{|W_{k+1} - \widetilde{W}_{k+1}| \ge 3\}} \\
&\qquad + \Bigg(|W_{k+1} - \widetilde{W}_{k+1}| + \frac{6}{k}|W_{k+1} - \widetilde{W}_{k+1}| \\
&\qquad\qquad + c\sum_{i=1}^{s} \frac{W_{k+1}^2(i) + \widetilde{W}_{k+1}^2(i)}{k^2}\Bigg) \cdot \mathbf{1}_{\{|W_{k+1} - \widetilde{W}_{k+1}| \le 2\}} \\
&\quad \le |W_{k+1} - \widetilde{W}_{k+1}|\left(1 - \frac{2}{k+1}\right) + \frac{16}{k} \\
&\qquad + c\sum_{i=1}^{s} \frac{W_{k+1}^2(i) + \widetilde{W}_{k+1}^2(i)}{k^2} + \frac{Z_{k+1}}{\binom{k+1}{2}}.
\end{aligned}$$

Taking expectation and using Lemma 1 and the fact that $k \le n$, we obtain that

$$\mathbb{E}(|W_k - \widetilde{W}_k|) \le \left(1 - \frac{2}{k+1}\right)\mathbb{E}\big[|W_{k+1} - \widetilde{W}_{k+1}|\big] + c\left(\frac{k^2}{n^2} + \frac{1}{k}\right).$$

Dividing the previous inequality by $k(k-1)$ we obtain a recurrence formula that we iterate from k up to $a_n - 1$,

$$
\begin{aligned}
&\frac{1}{k(k-1)}\mathbb{E}(|W_k - \widetilde{W}_k|) \\
&\quad \le \frac{1}{k(k+1)}\left(\mathbb{E}(|W_{k+1} - \widetilde{W}_{k+1}|) + c\left(\frac{k^2}{n^2} + \frac{1}{k}\right)\right) \\
&\quad \le \frac{1}{a_n(a_n-1)}\mathbb{E}(|W_{a_n} - \widetilde{W}_{a_n}|) + c\sum_{j=k}^{a_n-1}\left(\frac{1}{n^2} + \frac{1}{j^3}\right) \\
&\quad \le \frac{1}{a_n(a_n-1)}\big(\mathbb{E}(|W_{a_n} - \mathbb{E}(W_{a_n})|) + |\mathbb{E}(W_{a_n}) - \mathbb{E}(\widetilde{W}_{a_n})| \\
&\qquad\qquad + \mathbb{E}(|\widetilde{W}_{a_n} - \mathbb{E}(\widetilde{W}_{a_n})|)\big) + c\left(\frac{a_n}{n^2} + \frac{1}{k^2}\right).
\end{aligned}
$$

Finally by Lemma 1,

$$
\begin{aligned}
\frac{1}{k(k-1)}\mathbb{E}(|W_k - \widetilde{W}_k|) &\le c\left(\frac{1}{a_n(a_n-1)}\left(\frac{a_n}{\sqrt{n}} + \frac{a_n^3}{n^2} + \frac{a_n}{n}\right) + \frac{a_n}{n^2} + \frac{1}{k^2}\right) \\
&\le c\left(\frac{1}{a_n\sqrt{n}} + \frac{a_n}{n^2} + \frac{1}{k^2}\right).
\end{aligned}
$$

This gives the second claim of the lemma. Using this claim and the fact that $1 \le k \le a_n \le n$ in (12) yields the first claim. □

LEMMA 3. *There is a constant $c < \infty$ such that for $r \le s$ it holds that*

$$
\mathbb{V}\big(W_k(r) - \widetilde{W}_k(r)\big) \le c \cdot \left(\frac{k^2}{a_n\sqrt{n}} + \frac{a_n k^2}{n^2} + \frac{k^3}{a_n n} + 1\right)
$$

for all $1 \le k < a_n$.

PROOF. We again write here as an abbreviation W_k, $\widetilde{W}_k$, ΔW_k, and $\Delta\widetilde{W}_k$ instead of $W_k(r)$, $\widetilde{W}_k(r)$, $\Delta W_k(r)$, and $\Delta\widetilde{W}_k(r)$, respectively.

Using (14) and (15) together with the fact that $|\Delta W_{k-1} - \Delta\widetilde{W}_{k-1}| \le 3$, we obtain

$$
\begin{aligned}
&\mathbb{V}(W_{k-1} - \widetilde{W}_{k-1}) \\
&\quad = \mathbb{V}(W_k - \widetilde{W}_k + \Delta W_{k-1} - \Delta\widetilde{W}_{k-1}) \\
&\quad \le \mathbb{V}(W_k - \widetilde{W}_k) \\
&\qquad + 2\mathbb{E}\big(\mathbb{E}\big[\big(W_k - \widetilde{W}_k - \mathbb{E}(W_k - \widetilde{W}_k)\big) \\
&\qquad\qquad \times \big(\Delta W_{k-1} - \Delta\widetilde{W}_{k-1} - \mathbb{E}(\Delta W_{k-1} - \Delta\widetilde{W}_{k-1})\big)|\mathcal{F}_k\big]\big)
\end{aligned}
$$

$$
\begin{aligned}
&+\mathbb{E}(\Delta W_{k-1}-\Delta \widetilde{W}_{k-1})^2\\
&=\Big(1-\frac{4}{k}\Big)\mathbb{V}(W_k-\widetilde{W}_k)\\
&\quad+2\mathbb{E}\Big(\big(W_k-\widetilde{W}_k-\mathbb{E}(W_k-\widetilde{W}_k)\big)\cdot\frac{Z_k-\mathbb{E}(Z_k)}{\binom{k}{2}}\Big)\\
&\quad+9\mathbb{P}(W_{k-1}\neq \Delta\widetilde{W}_{k-1}).
\end{aligned}
$$

Applying now the Cauchy–Schwarz inequality for the second term and Lemma 2 for the third term on the left-hand side of the inequality above, we obtain that for a constant $c<\infty$

$$
\begin{aligned}
&\mathbb{V}(W_{k-1}-\widetilde{W}_{k-1})\\
&\quad\le\Big(1-\frac{4}{k}\Big)\mathbb{V}(W_k-\widetilde{W}_k)+\frac{4}{k(k-1)}\big(\mathbb{V}(W_k-\widetilde{W}_k)\big)^{1/2}\cdot\big(\mathbb{V}(Z_k)\big)^{1/2}\\
&\qquad+c\Big(\frac{k}{a_n\sqrt{n}}+\frac{a_nk}{n^2}+\frac{1}{k}\Big).
\end{aligned}
\tag{16}
$$

Let us now look closer at the variance of Z_k. In order to bound it from above, it is sufficient to bound the terms of the form $\mathbb{V}(W_k(i)W_k(j))$, $1\le i,j\le r$; see the definition of Z_k in (13). Writing as an abbreviation W_k' and W_k'' for $W_k(i)$ and $W_k(j)$, respectively, we have that

$$
\begin{aligned}
\mathbb{V}\big(W_k(i)W_k(j)\big)&\le\mathbb{E}\big(W_k'W_k''-\mathbb{E}(W_k')\mathbb{E}(W_k'')\big)^2\\
&=\mathbb{E}\big(\big(W_k'-\mathbb{E}(W_k')\big)W_k''+\mathbb{E}(W_k')\big(W_k''-\mathbb{E}(W_k'')\big)\big)^2\\
&\le 2\mathbb{E}\big(\big(W_k'-\mathbb{E}(W_k')\big)^2(W_k'')^2+\big(\mathbb{E}(W_k')\big)^2\big(W_k''-\mathbb{E}(W_k'')\big)^2\big).
\end{aligned}
$$

Using the fact that $W_k''\le k$ and then Lemma 1, we obtain

$$
\begin{aligned}
&\mathbb{V}\big(W_k(i)W_k(j)\big)\\
&\quad\le 2\mathbb{E}\big(\big(W_k'-\mathbb{E}(W_k')\big)^2\big(W_k''-\mathbb{E}(W_k'')\big)W_k''\big)\\
&\qquad+2\mathbb{E}\big(\big(W_k'-\mathbb{E}(W_k')\big)^2\mathbb{E}(W_k'')W_k''\big)+2\big(\mathbb{E}(W_k')\big)^2\mathbb{V}(W_k'')\\
&\quad\le 2\mathbb{E}\big(\big(W_k'-\mathbb{E}(W_k')\big)^2\big(W_k''-\mathbb{E}(W_k'')\big)^2\big)\\
&\qquad+2\mathbb{E}\big(\big(W_k'-\mathbb{E}(W_k')\big)^2\big(W_k''-\mathbb{E}(W_k'')\big)\mathbb{E}(W_k'')\big)\\
&\qquad+2k\cdot\mathbb{E}\big(\big(W_k'-\mathbb{E}(W_k')\big)^2\mathbb{E}(W_k'')\big)+2\big(\mathbb{E}(W_k')\big)^2\mathbb{V}(W_k'')\\
&\quad\le 2\mathbb{E}\big(\big(W_k'-\mathbb{E}(W_k')\big)^2\big(W_k''-\mathbb{E}(W_k'')\big)^2\big)\\
&\qquad+4k\cdot\mathbb{E}\big(\big(W_k'-\mathbb{E}(W_k')\big)^2\mathbb{E}(W_k'')\big)+2\big(\mathbb{E}(W_k')\big)^2\mathbb{V}(W_k'')\\
&\quad\le 2\mathbb{E}\big(\big(W_k'-\mathbb{E}(W_k')\big)^2\big(W_k''-\mathbb{E}(W_k'')\big)^2\big)+c\frac{k^5}{n^2}
\end{aligned}
\tag{17}
$$

for a constant $c < \infty$. Moreover, using the formulas from Lemma 1,

$$\begin{aligned}
&\mathbb{E}((W'_k - \mathbb{E}(W'_k))^4) \\
&\quad = \mathbb{E}((W'_k)^4) - 4\mathbb{E}((W'_k)^3)\mathbb{E}(W'_k) + 6\mathbb{E}((W'_k)^2)(\mathbb{E}(W'_k))^2 \\
&\qquad - 4\mathbb{E}(W'_k)(\mathbb{E}(W'_k))^3 + (\mathbb{E}(W'_k))^4 \\
&\quad = (1 - 4 + 6 - 4 + 1)\left(\frac{n-k}{n}\right)^{i-1} \cdot \frac{k^2}{n} + O\left(\left(\frac{k^2}{n}\right)^3 + \left(\frac{k^2}{n}\right)^2 + \frac{k^2}{n}\right).
\end{aligned}$$

The leading terms cancel, and since $(\frac{k^2}{n})^2$ is dominated by either $(\frac{k^2}{n})^3$ or $\frac{k^2}{n}$ depending on $k \geq \sqrt{n}$ or $k < \sqrt{n}$, we obtain that

$$\mathbb{E}((W'_k - \mathbb{E}(W'_k))^4) = O\left(\frac{k^6}{n^3} + \frac{k^2}{n}\right) \tag{18}$$

and similarly for W''_k

$$\mathbb{E}((W''_k - \mathbb{E}(W''_k))^4) = O\left(\frac{k^6}{n^3} + \frac{k^2}{n}\right). \tag{19}$$

Using the Cauchy–Schwarz inequality and (18) and (19) in (17), we get that

$$\mathbb{V}(W_k(i)W_k(j)) \leq c\left(\frac{k^6}{n^3} + \frac{k^5}{n^2} + \frac{k^2}{n}\right)$$

and therefore

$$\mathbb{V}(Z_k) \leq c\left(\frac{k^5}{n^2} + \frac{k^2}{n}\right)$$

for some constant $c < \infty$. Plugging this into (16) we obtain that

$$\begin{aligned}
\mathbb{V}(W_{k-1} - \widetilde{W}_{k-1}) \leq{}& \left(1 - \frac{4}{k}\right)\mathbb{V}(W_k - \widetilde{W}_k) \\
&+ c(\mathbb{V}(W_k - \widetilde{W}_k))^{1/2} \cdot \left(\frac{\sqrt{k}}{n} + \frac{1}{k\sqrt{n}}\right) \\
&+ c\left(\frac{k}{a_n\sqrt{n}} + \frac{a_n k}{n^2} + \frac{1}{k}\right).
\end{aligned} \tag{20}$$

Observe that

$$\begin{aligned}
&c(\mathbb{V}(W_k - \widetilde{W}_k))^{1/2} \cdot \left(\frac{\sqrt{k}}{n} + \frac{1}{k\sqrt{n}}\right) \\
&\qquad \leq \begin{cases} \dfrac{\mathbb{V}(W_k - \widetilde{W}_k)}{k}, & \text{if } (\mathbb{V}(W_k - \widetilde{W}_k))^{1/2} \geq c\left(\dfrac{k^{3/2}}{n} + \dfrac{1}{\sqrt{n}}\right), \\ 2c^2\left(\dfrac{k^2}{n^2} + \dfrac{1}{kn}\right), & \text{else} \end{cases}
\end{aligned}$$

and therefore since $k \le a_n$, (20) becomes

$$\mathbb{V}(W_{k-1} - \widetilde{W}_{k-1}) \le \left(1 - \frac{3}{k}\right)\mathbb{V}(W_k - \widetilde{W}_k) + c\left(\frac{k}{a_n\sqrt{n}} + \frac{a_n k}{n^2} + \frac{1}{k}\right).$$

We now divide both sides by $\binom{k-1}{3}$ and iterate up to a_n. Since $\mathbb{V}(W_{a_n} - \widetilde{W}_{a_n}) \le c\frac{a_n^2}{n}$, we obtain by Lemma 1

$$\begin{aligned}\frac{1}{\binom{k-1}{3}}\mathbb{V}(W_{k-1} - \widetilde{W}_{k-1}) &\le c \cdot \sum_{j=k}^{a_n} \frac{1}{\binom{j}{3}}\left(\frac{j}{a_n\sqrt{n}} + \frac{a_n j}{n^2} + \frac{1}{j}\right) \\ &\quad + c \cdot \frac{1}{\binom{a_n}{3}} \cdot \frac{a_n^2}{n}\end{aligned}$$

and therefore

$$\begin{aligned}\mathbb{V}(W_{k-1} - \widetilde{W}_{k-1}) &\le c \cdot k^3\left(\frac{1}{ka_n\sqrt{n}} + \frac{a_n}{kn^2} + \frac{1}{k^3} + \frac{1}{a_n n}\right) \\ &\le c \cdot \left(\frac{k^2}{a_n\sqrt{n}} + \frac{a_n k^2}{n^2} + \frac{k^3}{a_n n} + 1\right).\end{aligned}$$

This is the claim. $\square$

4. Proof of the theorem. The proof that μ_r, the expected length of order r, is equal to $\frac{2}{r}$ for every $r \ge 1$ can be found in [3], Theorem 2.11 or in [7], Theorem 2.1. Another quick way to see this is by using Lemma 1,

$$\begin{aligned}\mathbb{E}(\mathcal{L}^{n,r}) &= \mathbb{E}\left(\sum_{k=2}^{n} W_k(r) X_k\right) \\ &= \sum_{k=2}^{n} \mathbb{E}(W_k(r))\mathbb{E}(X_k) \\ &= \sum_{k=2}^{n} \frac{(n-k)\cdots(n-k-r+2)}{(n-1)\cdots(n-r)} \cdot k(k-1) \cdot \frac{1}{\binom{k}{2}} \\ &= \frac{2}{(n-1)\cdots(n-r)} \cdot \sum_{j=1}^{n-r} j(j+1)\cdots(j+r-2).\end{aligned}$$

The claim follows now from the fact that $\sum_{j=1}^{n} j(j+1)\cdots(j+i) = \frac{1}{i+2}n(n+1)\cdots(n+i+1)$. The asymptotic normality of the total external branch length of the Kingman coalescent (case $s=1$) was proved in [12]. We will prove the theorem for $s \ge 2$.

For $1 \le r \le s$ we divide $L^{n,r}$ and the corresponding coupled quantity into parts. For $1 \le b_n < a_n \le n$, let

$$L^{n,r}_{a_n,b_n} := \sum_{b_n < k \le a_n} \frac{2}{k(k-1)} \cdot W_k \quad \text{and} \quad \widetilde{L}^{n,r}_{a_n,b_n} := \sum_{b_n < k \le a_n} \frac{2}{k(k-1)} \cdot \widetilde{W}_k \tag{21}$$

be the length of order r collected between the levels b_n and a_n in the coalescent tree and the corresponding quantity obtained from the coupling. Note that $L^{n,r}_{n,1} = L^{n,r}$ with $L^{n,r}$ defined in (6), and let similarly

$$\widetilde{L}^{n,r} := \widetilde{L}^{n,r}_{n,1}. \tag{22}$$

Using the coupling we will show that for $\varepsilon > 0$

$$\begin{aligned}\mathbb{P}\biggl(\sqrt{\frac{n}{\log n}} \cdot \bigl\| \bigl(L^{n,1} - \mathbb{E}(L^{n,1}), \dots, L^{n,s} - \mathbb{E}(L^{n,s})\bigr) \\ - \bigl(\widetilde{L}^{n,1} - \mathbb{E}(\widetilde{L}^{n,1}), \dots, \widetilde{L}^{n,s} - \mathbb{E}(\widetilde{L}^{n,s})\bigr)\bigr\| \ge \varepsilon\biggr) \to 0 \\ \text{as } n \to \infty.\end{aligned} \tag{23}$$

Once (23) has been proved, the claim of the theorem follows since the components of the second vector above are by construction independent and identically distributed and they converge weakly to the standard normal distribution as $n \to \infty$, as follows from the case $s = 1$ proved in [12].

The convergence in (23) is a direct consequence of the following result.

PROPOSITION 1. *For $L^{n,r}$ and $\widetilde{L}^{n,r}$ defined in (6) and (22), respectively, one has for all $1 \le r \le s$ and $\varepsilon > 0$,*

$$\mathbb{P}\biggl(\biggl|\sqrt{\frac{n}{\log n}} \cdot \bigl(L^{n,r} - \mathbb{E}(L^{n,r})\bigr) - \bigl(\widetilde{L}^{n,r} - \mathbb{E}(\widetilde{L}^{n,r})\bigr)\biggr| \ge \varepsilon\biggr) \to 0, \qquad \textit{as } n \to \infty.$$

PROOF. We have by the Cauchy–Schwarz inequality that

$$\begin{aligned}&\mathbb{V}\biggl(\sum_{b_n < k \le a_n} \frac{2}{k(k-1)} \cdot \bigl(W_k - \widetilde{W}_k - \bigl(\mathbb{E}(W_k) - \mathbb{E}(\widetilde{W}_k)\bigr)\bigr)\biggr) \\ &\quad = \sum_{b_n < k \le a_n} \sum_{b_n < l \le a_n} \frac{2}{k(k-1)} \frac{2}{l(l-1)} \cdot \mathbb{COV}(W_k - \widetilde{W}_k, W_l - \widetilde{W}_l) \\ &\quad \le \sum_{b_n < k \le a_n} \sum_{b_n < l \le a_n} \frac{2}{k(k-1)} \frac{2}{l(l-1)} \cdot \mathbb{V}(W_k - \widetilde{W}_k)^{1/2} \mathbb{V}(W_l - \widetilde{W}_l)^{1/2} \\ &\quad = \biggl(\sum_{b_n < k \le a_n} \frac{2}{k(k-1)} \cdot \mathbb{V}(W_k - \widetilde{W}_k)^{1/2}\biggr)^2.\end{aligned}$$

Using Lemma 3 we obtain

$$\begin{aligned}
&\mathbb{V}\Bigg(\sum_{b_n<k\le a_n}\frac{2}{k(k-1)}\cdot\big(W_k-\widetilde{W}_k-\big(\mathbb{E}(W_k)-\mathbb{E}(\widetilde{W}_k)\big)\big)\Bigg)\\
&\quad\le c\Bigg(\sum_{b_n<k\le a_n}\frac{1}{k^2}\cdot\bigg(\frac{k}{\sqrt{a_n}n^{1/4}}+\frac{\sqrt{a_n}k}{n}+\frac{k^{3/2}}{\sqrt{a_n n}}+1\bigg)\Bigg)^2\\
&\quad\le c\bigg(\frac{1}{\sqrt{a_n}n^{1/4}}\log\frac{a_n}{b_n}+\frac{\sqrt{a_n}}{n}\log\frac{a_n}{b_n}+\frac{1}{\sqrt{n}}+\frac{1}{b_n}\bigg)^2\\
&\quad\le c\bigg(\bigg(\frac{1}{a_n\sqrt{n}}+\frac{a_n}{n^2}\bigg)\log^2\frac{a_n}{b_n}+\frac{1}{n}+\frac{1}{b_n^2}\bigg).
\end{aligned}\tag{24}$$

In order to show that the claim holds, we consider three regions in the coalescent tree, namely between level n and level $\frac{n}{(\log n)^2}$, between level $\frac{n}{(\log n)^2}$ and level $n^{1/2}$, and finally between level $n^{1/2}$ and level 1, and write the lengths $L_{n,1}^{n,r}$ and $\widetilde{L}_{n,1}^{n,r}$ as sums of the lengths gathered in these three regions.

For the first region, let

$$a_n=n \quad\text{and}\quad b_n=\frac{n}{(\log n)^2}.$$

We obtain from (24) and Chebyshev's inequality that

$$L_{a_n,b_n}^{n,r}-\mathbb{E}\big(L_{a_n,b_n}^{n,r}\big)=\widetilde{L}_{a_n,b_n}^{n,r}-\mathbb{E}\big(\widetilde{L}_{a_n,b_n}^{n,r}\big)+O_P\bigg(\frac{\log\log n}{\sqrt{n}}\bigg)$$

and therefore

$$\sqrt{\frac{n}{\log n}}\big(\big(L_{a_n,b_n}^{n,r}-\mathbb{E}\big(L_{a_n,b_n}^{n,r}\big)\big)-\big(\widetilde{L}_{a_n,b_n}^{n,r}-\mathbb{E}\big(\widetilde{L}_{a_n,b_n}^{n,r}\big)\big)\big)\to 0\tag{25}$$

in probability as $n\to\infty$.

The second region we consider is the one between the levels a_n and b_n with

$$a_n=\frac{n}{(\log n)^2}\quad\text{and}\quad b_n=n^{1/2}.$$

We put together the coupling for the two regions by taking the starting distribution for the second region to be the distribution of the chain at the end of the first region. Again from (24) we get that

$$L_{a_n,b_n}^{n,r}-\mathbb{E}\big(L_{a_n,b_n}^{n,r}\big)=\widetilde{L}_{a_n,b_n}^{n,r}-\mathbb{E}\big(\widetilde{L}_{a_n,b_n}^{n,r}\big)+O_P\bigg(\frac{1}{\sqrt{n}}\bigg)$$

and therefore as $n\to\infty$ in probability

$$\sqrt{\frac{n}{\log n}}\big(\big(L_{a_n,b_n}^{n,r}-\mathbb{E}\big(L_{a_n,b_n}^{n,r}\big)\big)-\big(\widetilde{L}_{a_n,b_n}^{n,r}-\mathbb{E}\big(\widetilde{L}_{a_n,b_n}^{n,r}\big)\big)\big)\to 0.\tag{26}$$

For the region in the coalescent between the levels $n^{1/2}$ and 1, we claim that

$$\mathbb{E}\left(\sqrt{\frac{n}{\log n}} \cdot L^{n,r}_{n^{1/2},1}\right) \to 0 \tag{27}$$

and

$$\mathbb{E}\left(\sqrt{\frac{n}{\log n}} \cdot \widetilde{L}^{n,r}_{n^{1/2},1}\right) \to 0 \tag{28}$$

as $n \to \infty$. The second claim follows directly from Proposition 3 in [12], whereas for (27) we get similarly using Lemma 1 that

$$\mathbb{E}(L^{n,r}_{a_n,b_n}) = \mathbb{E}\left(\sum_{b_n<k\le a_n} W_k \cdot \frac{2}{k(k-1)}\right) \tag{29}$$

$$\le c \sum_{b_n<k\le a_n} \frac{k^2}{n} \cdot \frac{1}{k(k-1)} \le c \cdot \frac{a_n}{n}, \tag{30}$$

for some constant $c < \infty$. Therefore, setting $a_n = n^{1/2}$ and $b_n = 1$ in (29), we obtain our claim (27). Since both $\sqrt{\frac{n}{\log n}} \cdot L^{n,r}_{n^{1/2},1}$ and $\sqrt{\frac{n}{\log n}} \cdot \widetilde{L}^{n,r}_{n^{1/2},1}$ are positive random variables, it follows from (27) and (28), respectively, that

$$\sqrt{\frac{n}{\log n}} \cdot L^{n,r}_{n^{1/2},1} \to 0 \quad \text{and} \quad \sqrt{\frac{n}{\log n}} \cdot \widetilde{L}^{n,r}_{n^{1/2},1} \to 0 \tag{31}$$

in probability as $n \to \infty$.

Writing

$$L^{n,r} = L^{n,r}_{n,1} = L^{n,r}_{n,(n/(\log n)^2)} + L^{n,r}_{n/(\log n)^2,n^{1/2}} + L^{n,r}_{n^{1/2},1}$$

and

$$\widetilde{L}^{n,r} = \widetilde{L}^{n,r}_{n,1} = \widetilde{L}^{n,r}_{n,(n/(\log n)^2)} + \widetilde{L}^{n,r}_{n/(\log n)^2,n^{1/2}} + \widetilde{L}^{n,r}_{n^{1/2},1}$$

and using (25)–(28) and (31), we get the claim of the proposition, and therefore our theorem is proved. $\square$

Acknowledgements. We thank two anonymous referees for careful reading and suggestions that improved the quality of the paper.

REFERENCES

[1] BASDEVANT, A.-L. and GOLDSCHMIDT, C. (2008). Asymptotics of the allele frequency spectrum associated with the Bolthausen–Sznitman coalescent. *Electron. J. Probab.* **13** 486–512. MR2386740

[2] BERESTYCKI, J., BERESTYCKI, N. and SCHWEINSBERG, J. (2007). Beta-coalescents and continuous stable random trees. *Ann. Probab.* **35** 1835–1887. MR2349577

[3] BERESTYCKI, N. (2009). *Recent Progress in Coalescent Theory*. *Ensaios Matemáticos* [*Mathematical Surveys*] **16**. Sociedade Brasileira de Matemática, Rio de Janeiro. MR2574323

[4] CALIEBE, A., NEININGER, R., KRAWCZAK, M. and RÖSLER, U. (2007). On the length distribution of external branches in coalescent trees: Genetic diversity within species. *Theor. Popul. Biol.* **72** 245–252.

[5] DAHMER, I., KERSTING, G. and WAKOLBINGER, A. (2014). The total external branch length of Beta-coalescents. *Combin. Probab. Comput.* **23** 1010–1027. MR3265836

[6] DHERSIN, J.-S. and YUAN, L. (2012). Asympotic behavior of the total length of external branches for Beta-coalescents. Available at arXiv:1202.5859.

[7] DURRETT, R. (2008). *Probability Models for DNA Sequence Evolution*, 2nd ed. Springer, New York. MR2439767

[8] FREUND, F. and MÖHLE, M. (2009). On the time back to the most recent common ancestor and the external branch length of the Bolthausen–Sznitman coalescent. *Markov Process. Related Fields* **15** 387–416. MR2554368

[9] FU, X. Y. (1995). Statistical properties of segregating sites. *Theor. Popul. Biol.* **48** 172–197.

[10] FU, Y. X. and LI, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* **133** 693–709.

[11] GNEDIN, A., IKSANOV, A. and MÖHLE, M. (2008). On asymptotics of exchangeable coalescents with multiple collisions. *J. Appl. Probab.* **45** 1186–1195. MR2484170

[12] JANSON, S. and KERSTING, G. (2011). On the total external length of the Kingman coalescent. *Electron. J. Probab.* **16** 2203–2218. MR2861672

[13] KERSTING, G., PARDO, J. C. and SIRI-JÉGOUSSE, A. (2014). Total internal and external lengths of the Bolthausen–Sznitman coalescent. *J. Appl. Probab.* **51A** (Celebrating 50 Years of Applied Probability Trust). To appear.

[14] KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248. MR0671034

[15] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937

[16] MÖHLE, M. (2010). Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson–Dirichlet coalescent. *Stochastic Process. Appl.* **120** 2159–2173. MR2684740

[17] WAKELEY, J. (2008). *Coalescent Theory*: *An Introduction*. Roberts & Company, Greenwood Village, CO.

INSTITUT FÜR MATHEMATIK
GOETHE-UNIVERSITÄT FRANKFURT
ROBERT-MAYER-STR. 10
BOX 187
D-60325 FRANKFURT AM MAIN
GERMANY
E-MAIL: dahmer@math.uni-frankfurt.de
kersting@math.uni-frankfurt.de