High Order Formation and Evolution of Hornerin in Primates

Vanessa Romero1,2, Hirofumi Nakaoka1,2, Kazuyoshi Hosomichi3, and Ituro Inoue1,2,*

1Department of Genetics, School of Life Sciences, SOKENDAI (Graduate University for Advanced Studies), Mishima, Japan
2Division of Human Genetics, National Institute of Genetics, Mishima, Japan
3Department of Bioinformatics and Genomics, Graduate School of Medical Sciences, Kanazawa University, Japan

*Corresponding author: E-mail: tinoue@nig.ac.jp.

Accepted: September 25, 2018

Abstract

Genomic duplication or loss can accelerate evolution because the number of repeats could affect molecular pathways and phenotypes. We have previously reported that the repeated region of filaggrin (FLG), a crucial component of the outer layers of mammalian skin, had high levels of nucleotide diversity with species-specific divergence and expansion and that it evolved under the birth-and-death model. We focused on hornerin (HRNR), a member of the same gene family that harbor similar tandem repeats as FLG, and examined the formation process of repeated regions and the evolutionary model that best fit the HRNR repeated region in the crab-eating macaque (Macaca fascicularis), orangutan (Pongo abelii), gorilla (Gorilla gorilla), and chimpanzee (Pan troglodytes) and compared them with the human (Homo sapiens) sequence. Paar et al. (2011) and Takaishi et al. (2005) have different theories as to the formation of the repeated region of HRNR; both groups share the longest repeat length of 1,404 bp (quartic or longest unit), but they differed in the process. We identified the formation described by Paar et al. [((“39 bp (primary) \times 9” \times 2 (secondary)) \times 2 (tertiary)) \times 5 (quartic)] to be conserved in all species except the crab-eating macaque. We detected high nucleotide diversities between the longest repeats, which fits the birth-and-death model. We concluded that the high order repeat formation of HRNR was conserved in primates except the crab-eating macaque. As previously identified in FLG, the longest repeats have high levels of nucleotide diversity, which could contribute to phenotypic differences between closely related species.

Key words: birth-and-death model, epidermal differentiation complex, hornerin, HRNR, primate, S100 fused type protein.

Introduction

The epidermal differentiation complex (EDC) locus, which is involved in immunity, perception, and epithelia, was identified as one of the most rapidly evolving loci in the human (Homo sapiens) genome. EDC on chromosome 1q21 spans ~1.6 Mb comprising 65 genes represented by four gene families, that is, S100 fused type protein (SFTP), late cornified envelope, small proline repeat-rich, and S100-domain genes. SFTP family members locate in a clustered manner, share a tandem repeat structure on the third exon, and have a similar function in the cornified epithelium of skin and, in conjunction with other members of the SFTP, reinforces the envelopes of the keratinizing epithelia (Wu et al. 2009; Henry et al. 2012; Kypriotou et al. 2012). We previously reported that the repeated region of filaggrin, an important component of the outer layers of mammalian skin, has high levels of nucleotide diversity with species-specific divergence and expansion, and evolves under the birth-and-death model (Romero et al. 2017). There is a possibility that all the SFTP family members commonly evolve under the birth-and-death model because their repeated structure rendered the selective advantage and resulted the rapid evolution. Here, we focus on the hornerin gene (HRNR), a member of SFTP, which has not been well characterized thus far.

HRNR contains a tandem repeat structure like filaggrin and is expressed in regenerating and normal skin mostly observed on the head, trunk, legs, hands, and feet. HRNR is a component of the cornified epithelium of skin and, in conjunction with other members of the SFTP, reinforces the envelopes of the keratinizing epithelia (Henry et al. 2012). The formation of the repeat region of HRNR is controversial. Takaishi et al. (2005) described HRNR formation as starting with 117 bp, which was later amplified 4 times (468 bp), tripled to form 1,404 bp in tandem, and then further amplified 6-fold to form the repeated region of ~8,424 bp (fig. 1A). On the contrary, Paar et al. (2011) described HRNR formation using the Global Repeat Map algorithm as follows: 39 bp, a “primary repeat unit” sequence, was amplified by nine to form a “secondary...
We performed the multiple nucleotide sequence alignment using the profile alignment in ClustalW implemented in MEGA 6.06 (Tamura et al. 2011) using the longest human unit described by Takaishi et al. (2005) as a reference (supplementary table S1, Supplementary Material online). We then constructed neighbor-joining trees and maximum likelihood trees. Neighbor-joining trees were constructed by considering pairwise deletions for an average of 1,200 nucleotides, proportional nucleotide differences (p-distance), and 1,000 bootstrap resampling. We compared 24 DNA/Protein models in MEGA 6.06 and the best-fit model was Kimura's two-parameter substitution model assuming that the rate of

**Materials and Methods**

**HRNR Database Sequences**

We obtained full-length *HRNR* DNA sequences from the National Center for Biotechnology Information (NCBI) gene database (http://www.ncbi.nlm.nih.gov/gene; last accessed October 11, 2018) for the following species: *H. sapiens* (NC 000001.11), *P. troglodytes* (NC 006468.3), *G. gorilla* (NC 018424.1), *P. abelii* (NC 012591.1), and *M. fascicularis* (NC 022272.1).

**Identification of *HRNR* Repeated Units in Primates**

We used the amino acid description of the longest unit given by Takaishi et al. (2005), which was subdivided into three types of the second longest units (A, B, and C). Each was further divided into four initial 117-bp units (I–IV), and the rest of the nucleotide sequence repeats were detected in humans using the Align Sequences Nucleotide Basic Local Alignment Search Tool (BLAST) with an identity >80% (Zhang et al. 2000) (fig. 1 and supplementary fig. 1A, Supplementary Material online). As described by Takaishi et al. (2005), the 6th longest unit does not contain the type-C 2nd longest unit (fig. 2). Next, using BLAST (Zhang et al. 2000), we compared the repeated sequences from humans with those in the crab-eating macaque, orangutan, gorilla, and chimpanzee. Finally, the matched sequences were reanalyzed by BLAST and then manually curated (figs. 1 and 2, and supplementary fig. 1A, Supplementary Material online). In the case of the crab-eating macaque, we also acquired a dot-matrix plot that compared each initial 117-bp units of humans with the 4th longest unit of the crab-eating macaque (Zhang et al. 2000) (supplementary fig. S2, Supplementary Material online). Each of the identified longest units, second longest, and initial 117-bp units were considered as independent units for the subsequent analyzes. Additionally, we used the 39-bp sequence of the primary unit described by Paar et al. (2011) and detected the primary units for human, chimpanzee, gorilla, orangutan, and crab-eating macaque and then reconstructed the secondary, tertiary, and quartic units using BLAST (supplementary fig. 1B, Supplementary Material online).

The multiple alignment viewer Mview (Brown et al. 1998) allowed us to observe the alignment location between each of the initial 117-bp units of humans with the 4th longest unit of the crab-eating macaque, orangutan, and chimpanzee and the 2nd longest unit of the gorilla (supplementary fig. S3, Supplementary Material online).

The percentage of similarity by a pairwise comparison between the longest units was conducted using the percent identity matrix included in the web services of Clustal Omega at the European Molecular Biology Laboratory-European Bioinformatics Institute (McWilliam 2013) (http://www.ebi.ac.uk/Tools/msa/clustalo/; last accessed October 11, 2018).
substitution for each site follows a Gamma distribution with a BIC of 16041.42 (supplementary table S2, Supplementary Material). Maximum likelihood trees were constructed using the following predefined parameters: partial deletions, Kimura’s two-parameter substitution model assuming that the rate of substitution for each site follows a Gamma distribution, the nearest-neighbor-interchange heuristic method, and 1,000 bootstrap resampling. A cutoff value of 50% was set for the condensed tree (Romero et al. 2017).

Estimation of Polymorphic/Variant Sites, Nucleotide Diversity, and Ratio of Synonymous and Nonsynonymous Sites Using the DNAsp5 Program

The number of nonsynonymous and synonymous substitutions between repeats within a species was calculated using DNA polymorphism and estimating the synonymous and nonsynonymous change options in DNAsp5 (Librado and Rozas 2009). DNAsp5 estimated the nucleotide diversity (θ), which is the average number of nucleotide differences per site.
between sequences. DNAsp5 also calculated the ratio of the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon (Ka) to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon (Ks) (Hu and Banzhaf 2008; Librado and Rozas 2009). We estimated the Ka/Ks ratio for each pair of within-species repeats including gaps using the program DNAsp5. The level of purifying selection was then calculated as \((1 - Ka/Ks) \times 100\).

Inference of Gene Duplication and Loss in the Species Tree

We used the NOTUNG program to reconcile the species tree that had the best fit for duplication, transfer, loss, and incomplete lineage sorting events from the maximum likelihood tree described earlier (Durand et al. 2006; Vernot et al. 2008; Stolzer et al. 2012).

Results

**HRNR Sequences in Primates**

The hornerin gene HRNR is localized on chromosome 1 of the crab-eating macaque, orangutan, gorilla, chimpanzee, and human. HRNR comprises three exons with a conserved exon/intron composition between the mouse and human. The repeat region is embedded in the third exon and does not include introns (Henry et al. 2012; Kypriotou et al. 2012). The DNA sequences of HRNR were obtained from the NCBI gene database for human, chimpanzee, gorilla, orangutan, and crab-eating macaque. The length of the longest repeat formed was 1,404 bp and shared both formations proposed by Paar et al. (2011) and Takaishi et al. (2005) (fig. 1). In the subsequent analyses, each of the repeats was considered as an independent unit.

The number of the longest units was six for human, chimpanzee, orangutan, and crab-eating macaque and four for gorilla (fig. 2). Each longest unit comprising second longest units (A, B, and C) varied from 981 to 1,443 bp, and the last longest unit lacking the second longest unit C ranged from 936 to 945 bp (fig. 2). The similarity between the longest units within a species ranged from 75.8% for the crab-eating macaque to 98.3% for human (supplementary table S3, Supplementary Material online).

**Phylogenetic Analysis of Longest Units across Species**

Based on the results of the multiple sequence alignment of the longest unit (supplementary table S1, Supplementary Material online), we constructed phylogenetic trees using the neighbor-joining and maximum likelihood methods with a cutoff value of 50% due to the high similarity of the repeats, as described in Materials and Methods. The crab-eating macaque was used as an outgroup (fig. 3 and supplementary fig. S4A, Supplementary Material online).

Both methods gave similar phylogenetic trees (fig. 3 and supplementary fig. S4A, Supplementary Material online). In the maximum likelihood tree, the longest units, from 2 to 4 for crab-eating macaque (cluster “α”), from 2 to 5 for orangutan (cluster “δ-3”), and 3 and 5 units for human (cluster “ε”) clustered within the species. Some of gorilla and chimpanzee repeats clustered across species as follows: chimpanzee 6th longest unit and gorilla 4th longest-unit were in cluster “δ.” Chimpanzee 4th longest-unit and orangutan 1st longest-unit were in cluster “δ.” Finally, human 4th longest-unit and 6th longest-unit, orangutan 6th longest-unit, and crab-eating macaque 1st longest-unit, 5th longest-unit, and 6th longest-unit did not cluster with any repeat.

**Fig. 3.**—Maximum likelihood tree reconstruction using hornerin longest-units in these primates and the following parameters: partial deletions, Kimura’s two-parameter substitution model assuming that the rate of substitution for each site follows Gamma distribution, the nearest-neighbor-interchange heuristic method, and 1,000 bootstrap resampling. A cutoff value for the condensed tree was set at 50%. Crab-eating macaque was used as an outgroup. The longest-units that clustered within the species are found in clusters “α,” “δ-3,” and “ε.” Some of gorilla and chimpanzee repeats clustered across species in clusters “β” and “γ.” The rest of the repeats formed a big cluster “δ” across species, which was further subdivided into clusters from “δ-1” to “δ-4.” Chimpanzee 4th longest-unit and orangutan 1st longest-unit were in cluster “δ.” Finally, human 4th longest-unit and 6th longest-unit, orangutan 6th longest-unit, and crab-eating macaque 1st longest-unit, 5th longest-unit, and 6th longest-unit did not cluster with any repeat.
repeats from 2 to 5 (“δ-3”), and gorilla, chimpanzee, and human 1st longest units (“δ-4”) (fig. 3). Chimpanzee 4th longest unit and orangutan 1st longest unit were in cluster “δ-5.” Finally, human 4th and 6th longest units, orangutan 6th longest unit, and crab-eating macaque 1st, 5th, and 6th longest units did not cluster with any repeat. The neighboring joining tree further grouped the human 4th longest unit into cluster “δ-6,” the crab-eating macaque 1st longest unit into cluster “δ-7,” and the orangutan 6th longest unit and crab-eating macaque 5th and 6th longest units into cluster “β” (supplementary fig. S4A, Supplementary Material online).

Additionally, we constructed a species-gene tree and identified the duplications and losses in the phylogeny. Clusters “α,” “δ-3,” and “δ-5” were the results of duplications within the crab-eating macaque, orangutan, and human, respectively. Five losses within cluster “δ” and six losses on the remaining branches were identified between the orangutan, gorilla, chimpanzee, and human (supplementary fig. S4B, Supplementary Material online).

**HRNR Formation in Primates**

**Primary Units (39 bp) Described by Paar et al. (2011)**

The proposed formation by Paar et al. (2005) was \((117 \text{bp} \times 4) \times 3 \text{ subunits} \times 6 \text{ units}\) (fig. 1A), whereas the formation by Paar et al. (2011) was \((117 \text{bp} \text{ primary} \times 9) \times 2 \text{ (secondary)} \times 2 \text{ (tertiary)} \times 5 \text{ (quartic)}\) (fig. 1B). We examined both possibilities and clarified which structure was detectable in primates.

We started with the primary unit (39 bp) sequence and estimated the primary units for all primates using BLAST (fig. 4A and B). The phylogenetics of the primary units divided them into nine clusters (order, 1–9) in each species tree by Paar et al. (2011). This pattern was observed for human, chimpanzee, gorilla, and orangutan (fig. 4, Supplementary Material online). This pattern was observed for human, chimpanzee, gorilla, and orangutan (fig. 4, Supplementary Material online). We concluded the proposed pattern by Takaishi et al. (2005) was not detected; we observed the following overall clusters: A-I, A-IV, B-III, and C-II (“Group-1st”); A-II, B-I, B-IV, and C-III (“Group-2nd”); and A-III, B-II, C-I, and C-IV (“Group-3rd”), with the exception of a few initial 117-bp units in human B-III, chimpanzee A-I and B-IV, B-I to II, and B-I to III, gorilla B-I and II, and C-I, orangutan A-III and B-III, B-I to IV, and C-I and III, and all orangutan C-II and IV (fig. 5 and supplementary figs. S3 and S5A–C, Supplementary Material online). The phylogenetic clustering implies the following sequential pattern: “Group-1st,” “Group-2nd,” and “Group-3rd,” which fit the length of the secondary unit (351 bp) described by Paar et al. (2011) (figs. 1B and 5).

Next, we used the second longest units to construct a phylogenetic tree (figs. 1A and 2). Indeed, the tree clustered together each of the second longest units A to C (supplementary fig. S7, Supplementary Material online); however, the phylogeny of the initial 117-bp unit was not consistent with the duplication pattern described by Takaishi et al. (2005). We concluded the proposed pattern by Takaishi et al. unlikely (fig. 5 and supplementary fig. S6A–C, Supplementary Material online).

**Secondary Unit (351 bp) by Paar et al. (2011)**

We reconstructed the secondary units organized by nine repeats of the primary unit as described before, with the exception of the crab-eating macaque, in which the order of the primary units was not sequential (fig. 2). We constructed a phylogenetic tree for the secondary units across species (supplementary fig. S8, Supplementary Material online). The phylogeny for the secondary units gathered an n01 type, n02 type, and the last units for gorilla (except for w04-v01-n01 and w04-v02-n01), orangutan, chimpanzee, and human secondary units (supplementary fig. S8, Supplementary Material online).

**Tertiary Units (702 bp) by Paar et al. (2011)**

We next focused on tertiary units proposed by Paar et al. (2011). We used the secondary units to reconstruct the tertiary units (fig. 2). Paar et al. (2011) divided tertiary units into “v01” and “v02” types. We constructed a phylogenetic tree...
for the tertiary units across species and identified clusters for the v01 type, v02 type, and the last unit for chimpanzee (except for w03-v01), gorilla (except for w01-v02, w02-v01, w02-v02, and w03-v01), orangutan, and human (supplementary fig. S9, Supplementary Material online). The tertiary units were clustered similar to the phylogeny of the secondary units. In the phylogeny of the secondary and tertiary units, we also observed duplications within species, more commonly in orangutan and human (supplementary figs. S7 and S8, Supplementary Material online). Taking into consideration our results from the analyses of both Takaishi et al. (2005) and Paar et al. (2011), we concluded that the secondary units duplicated to form the tertiary units and then again duplicated to form quartic units. According to the description by Paar et al. (2011), this organization was only detected in human but not in chimpanzee. However, our analyses detected the organization to be conserved in chimpanzee, gorilla, and orangutan by the longest unit clustering found in the phylogenetic analysis (fig. 3) using the latest version of the NCBI gene database.

**Quartic Units (1404 bp) by Paar et al. (2011) or the Longest Unit by Takaishi et al. (2005)**

As for the quartic unit repeat, Takaishi et al. (2005) described six longest units and Paar et al. (2011) described five. Because the last unit contains only the second longest units of type A and B, it was not detected by the Global Repeat Maker algorithm developed by Paar et al. (2011) (fig. 1B). Multiple alignment and phylogenetic analyses for the longest unit were described in the previous section.

**Nucleotide Diversity in the Longest Units of Hornerin**

The concerted and birth-and-death models have been proposed to explain the nondivergent evolutionary pattern found in repeated gene families. In the concerted model, variations in one repeat spread to the others by unequal crossover or gene

![Diagram](https://example.com/diagram.png)
conversion, maintaining homogeneity among the repeats (Nei et al. 2000). In the birth-and-death model, new repeat genes diversify by silent nucleotide substitutions, which, with enough divergence time, can lead to lineage-specific expansions (Nei et al. 2000). The main difference between these two models is the high levels of intragenic nucleotide diversity that are only found in the birth-and-death model (Nei et al. 2000).

We calculated the total number of sites showing variation and nucleotide diversity (π) between the longest units of HRNR in each primate using the DNAsp5 program, as described in Materials and Methods. The estimated π of human, chimpanzee, gorilla, orangutan, and crab-eating macaque was $1.1 \times 10^{-2}$, $1.4 \times 10^{-2}$, $1.6 \times 10^{-2}$, $1.4 \times 10^{-2}$, and $1.7 \times 10^{-2}$, respectively (table 1). We then compared π of HRNR with that of the polyubiquitin gene and filaggrin repeats (Romero et al. 2017), and rDNA genes, which are representative multigene families that have evolved under birth-and-death evolution and concerted evolution, respectively (Nei et al. 2000; Ganley and Kobayashi 2007). The π of HRNR was comparable with that of polyubiquitin genes (range $= 8.9 \times 10^{-2}$ to $2.0 \times 10^{-1}$) and filaggrin repeats (range $= 3.2 \times 10^{-2}$ to $7.7 \times 10^{-2}$), and was much larger than those of rDNA genes (range $= 1.0 \times 10^{-5}$ to $1.8 \times 10^{-4}$) (Nei et al. 2000; Ganley and Kobayashi 2007). This suggests that HRNR longest units have evolved under the birth-and-death model.

Selection within the Repeats of Each Species

The birth-and-death model maintains repeat similarity by purifying selection. Under purifying selection, the number of nonsynonymous variations in a gene is expected to be smaller than the number of synonymous variations. Purifying selection can be measured by making pairwise comparisons $[(1 - Ka/Ks) \times 100]$, which indicate the percentage of synonymous mutations (Nei 2007).
We found that the average percentage of synonymous substitutions for human, chimpanzee, gorilla, orangutan, and crab-eating macaque was 46.30%, 36.84%, 44.16%, 29.05%, and 49.47%, respectively (table 1). Although nucleotide substitutions were generally synonymous, we found three pairs of slightly higher nonsynonymous variations in the longest units within species; the 3rd and 2nd longest units, 3rd and 4th longest units, and 3rd and 5th longest units. None of the pairs of repeats in the crab-eating macaque, gorilla, chimpanzee, or human had Ka/Ks > 1 (supplementary table S4, Supplementary Material online). We concluded that the high synonymous variation detected in the longest units fits the birth-and-death model.

**Discussion**

Hornerin has a unique and complex duplicate formation that is different from other SFTPs. The hornerin repeat formations proposed by Takaishi et al. (2005) [((117 bp × 4 initial) × 3 subunits] × 6 units) and by Paar et al. (2011) [“39 bp × 9 primary” × 2 secondary] × 2 tertiary) × 5 quartic] were examined and compared among several primates (fig. 1).

In the repeat organization by Paar et al. (2011), 39 bp of the primary unit was detected in chimpanzee, gorilla, orangutan, and crab-eating macaque (fig. 2). The phylogeny of the primary units per species divided them into nine clusters in chimpanzee, gorilla, and orangutan but not the crab-eating macaque (fig. 4 and supplementary fig. S5, Supplementary Material online). We did not observe the expected phylogeny for the initial 117-bp unit with a 4-fold duplication pattern as described by Takaishi et al. (2005) (fig. 5). The phylogeny for most human, chimpanzee, gorilla, and orangutan secondary units clustered in n01, n02, and the last units and the tertiary units clustered in v01, v02, and the last units (supplementary fig. S8, Supplementary Material online). We observed clustering within human and orangutan, suggesting unique duplications in these species (supplementary fig. S8, Supplementary Material online). We concluded that the primary unit is conserved and detectable in all primates but the duplication clustering order from 1 to 9 was not found in the crab-eating macaque (supplementary fig. S6–G, Supplementary Material online). We confirmed that in all primates except the crab-eating macaque the formation model started with the primary units which duplicated to make the secondary units, and the secondary units duplicated twice again to form the tertiary structure, as described by Paar et al. (2011) (fig. 18).

The longest units or quartic units, common to both Takaishi et al. (2005) and Paar et al. (2011), were used to understand hornerin evolution. Two models have been proposed to explain the nondivergent evolutionary pattern found in repeated gene families; concerted and birth-and-death (Nei et al. 2000). In the concerted model, the low level of diversity of repeats within a species is the results from unequal crossover or gene conversion, and the expected phylogeny clusters together the more similar repeats within a species. In the birth-and-death model, there is a high level of diversity between repeats, and genes have a close interspecies pattern in the phylogeny (Nei et al. 2000). Similar to filaggrin, hornerin units have high synonymous nucleotide diversity together with various lineage-specific expansions that fit the birth-and-death model (Nei et al. 2000; Nei and Rooney 2005; Nei 2007; Eirín-López et al. 2012; Sabbagh et al. 2013) (table 1 and supplementary tables S3 and S4, Supplementary Material online). The birth-and-death model has been used to explain most of multigene families evolution. One example is the Odorant-Binding Protein (OBP) family of Drosophila species (Vieira et al. 2007). All members from the OBP family show tandem repeat duplications evolving under purifying selection with different functional constraints per gene and a progressive nucleotide divergence, and likely rapidly affecting the sensitivity or specificity in detecting odors under different environments (Eirín-López et al. 2012). Another example is the fatty acid reductase (FAR) multigene family. The number of FAR genes per genome can vary greatly between organisms and phylogenetic relationships of its representatives show a pattern of between-species gene clustering, including some relative long branches. Most of the expansions occurred in plants and insects compared with other taxa and is the underlying cause for their ability to synthesize and utilize a wide variety of fatty alcohol-based or derived compounds for a number of highly specialized functions (Eirín-López et al. 2012). As for the OBP and FAR families, the gene divergence found in different species result in rapid functional changes. It has previously been reported that gene-associated tandem repeats act as an accelerator of evolution by generating variation in structure and functionality (Fondon and Garner 2004). Under the birth-and-death model, duplicates vary by silent nucleotide variations, which, with enough divergence time, can lead to lineage-specific expansions, especially...
observed in multigene families evolution. The similar evolutionary pattern for hornerin and filaggrin could be an indication that members of the SFTP family also evolve under the birth-and-death model (Vieira et al. 2007). The possibility of all the members of SFTP evolving under this model could explain the advantage of their repeated structure resulting in the rapid evolution required by genes from skin.

Hornerin variation within a species has not been reported; however, the biological importance of tandem repeats was mentioned by Paar et al. (2011) as a rapidly evolving type of DNA sequences that could contribute to phenotypic differences, even between closely related species such as humans and chimpanzees. The Global Repeat Map of Paar et al. (2011) failed to find the high order repeat of human in chimpanzee; however, our analysis detected the 39-bp length to be conserved in all primates except the crab-eating macaque, probably due to differences in the database version. Our analysis also detected unique duplications for all primates, especially for human and orangutan in the secondary and tertiary phylogeny. Neuroblastoma break-point family (NBFP) copy number variability is a dramatic example of high order repeat in human and chimpanzee (Paar et al. 2011). The NBFP repeat is related to the evolutionary level of higher primates and the high order repeat pattern shows a discontinuous jump in the evolutionary step from 48 monomers in chimpanzee to 165 monomers in human, possibly related to a regulatory function of high order repeats (Paar et al. 2011). Indeed, there was no change in the number of hornerin copies between chimpanzees and humans; however, the duplications found in the phylogeny could possibly contribute to a difference in phenotype. Although the functional significance of copy number variation of HRNR repeats in primates requires further studies, we suggest that, like NBFP and FLG, HRNR high repeat order pattern might contribute to a different phenotype.

Supplementary Material
Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments
The authors would like to thank Enago (www.enago.jp) for the English language review. The study was funded by the basal running budget from the National Institute of Genetics, Japan.

Literature Cited
Brown N, Leroy C, Sander C. 1998. MView: a web-compatible database search or multiple alignment viewer. Bioinformatics 14(4):380–381.
Durand D, Hallbööksson BV, Vernot B. 2006. A hybrid macro-macroevolutionary approach to gene tree reconstruction. J Comput Biol. 13(2):320–335.
Eirín-López J, Reborondos L, Rooney A, Rozas J. 2012. The birth-and-death evolution of multigene families revisited. Genome Dyn. 7:170–196.
Fondón JW, Garner HR. 2004. Molecular origins of rapid and continuous morphological evolution. Proc Natl Acad Sci U S A. 101(52):18058–18063.
Ganley ARD, Kobayashi T. 2007. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole genome shotgun sequence data. Genome Res. 17(2):184–191.
Henry J, et al. 2012. Update on the epidermal differentiation complex. Front Biol. 7:1517–1532.
Hu T, Banzhaf W. 2008. Nonsynonymous to synonymous substitution ratio ka/kds: measurement for rate of evolution in evolutionary computation. In: Rudolph G, et al. eds, R. Parallel Problem Solving from Nature – PPSP X. Dortmund: Springer. P. 448–457.
Kypriotou M, Huber M, Hohl D. 2012. The human epidermal differentiation complex: cornified envelope precursors, S100 proteins and the ‘fused genes’ family. Exp Dermatol. 21(9):643–649.
Librado P, Rozas J. 2009. DNAp v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11):1451–1452.
McWilliam H. 2013. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. 41(Web Server issue):W597–W600.
Nei M. 2007. The new mutation theory of phenotypic evolution. Proc Natl Acad Sci U S A. 104(30):12235–12242.
Nei M, Rognes I, Piontkivska H. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. Proc Natl Acad Sci U S A. 97(20):10866–10871.
Nei M, Rooney A. 2005. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 39:121–152.
Paar V, Gluncic M, Rosandic M, Basar I, Vlahovic I. 2011. Intragenic higher order repeats in neuroblastoma breakpoint family genes distinguish humans from chimpanzees. Mol Bio Evol. 28(6):1877–1892.
Romero V, Hosomichi K, Naoka K, Shibata H, Inoue I. 2017. Structure and evolution of the filaggrin gene repeated region in primates. BMC Evol Biol. 17(1):10.
Sabbagh A, et al. 2013. Rapid birth-and-death evolution of the xenobiotic metabolizing NAT gene family in vertebrates with evidence of adaptive selection. BMC Evol Biol. 13:62.
Stolzer M, et al. 2012. Inference of duplications, losses, transfers, and incomplete lineage sorting with non-binary species trees. Bioinformatics 28(18):2409–2415.
Takahashi M, Makino T, Morohashi M, Huh NH. 2005. Identification of human hornerin and its expression in regenerating and psoriatic skin. J Biol Chem. 280(6):4696–4703.
Tamura K, et al. 2011. MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28(10):2731–2739.
Vernot B, Stolzer M, Goldman A, Durand D. 2008. Reconciliation with non-binary species trees. J Comput Biol. 15(8):981–1006.
Vieira FG, Sánchez-Gracia A, Rozas J. 2007. Comparative genomic analysis of the odorant-binding protein family in 12 Drosophila genomes: purifying selection and birth-and-death evolution. Genome Biol. 8(11):R235.
Wu Z, Hansmann B, Meyer-Hoffert U, Gläser R, Schröder J-M. 2009. Molecular identification and expression analysis of filaggrin-2, a member of the S100 fused-type protein family. PLoS One 4(4):e5227.
Wu Z, Meyer-Hoffert U, et al. 2009. Highly complex peptide aggregates of the S100 fused-type protein hornerin are present in human skin. J Invest Dermatol. 129(6):1446–1458.
Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol. 7(1-2):203–214.

Associate editor: Partha Majumder