Duplications in Corneous Beta Protein Genes and the Evolution of Gecko Adhesion

Tony Gamble1,*,†,‡

*Department of Biological Sciences, Marquette University, Milwaukee, WI 53201, USA; †Bell Museum of Natural History, University of Minnesota, Saint Paul, MN 55113, USA; ‡Milwaukee Public Museum, Milwaukee, WI 53233, USA

From the symposium “The path less traveled: Reciprocal illumination of gecko adhesion by unifying material science, biomechanics, ecology, and evolution” presented at the annual meeting of the Society of Integrative and Comparative Biology, January 3–7, 2019 at Tampa, Florida.

1E-mail: tgamble@geckoevolution.org

Synopsis  Corneous proteins are an important component of the tetrapod integument. Duplication and diversification of keratins and associated proteins are linked with the origin of most novel integumentary structures like mammalian hair, avian feathers, and scutes covering turtle shells. Accordingly, the loss of integumentary structures often coincides with the loss of genes encoding keratin and associated proteins. For example, many hair keratins in dolphins and whales have become pseudogenes. The adhesive setae of geckos and anoles are composed of both intermediate filament keratins (IF-keratins, formerly known as alpha-keratins) and corneous beta-proteins (CBPs, formerly known as beta-keratins) and recent whole genome assemblies of two gecko species and an anole uncovered duplications in seta-specific CBPs in each of these lineages. While anoles evolved adhesive toepads just once, there are two competing hypotheses about the origin(s) of digital adhesion in geckos involving either a single origin or multiple origins. Using data from three published gecko genomes, I examine CBP gene evolution in geckos and find support for a hypothesis where CBP gene duplications are associated with the repeated evolution of digital adhesion. Although these results are preliminary, I discuss how additional gecko genome assemblies, combined with phylogenies of keratin and associated protein genes and gene duplication models, can provide rigorous tests of several hypotheses related to gecko CBP evolution. This includes a taxon sampling strategy for sequencing and assembly of gecko genomes that could help resolve competing hypotheses surrounding the origin(s) of digital adhesion.

Introduction

Geckos are well known for their climbing abilities. Approximately 60% of gecko species have adhesive digits that facilitate climbing vertical surfaces, while the remaining gecko species lack adhesive toepads (Pianka and Vitt 2003; Gamble et al. 2012). Adhesion is mediated by subdigital setae, complex, hair-like projections of the epidermis, that mainly operate via Van der Waals forces (Autumn et al. 2002). Setae are hypothesized to have evolved from spinules, simple, microscopic projections found on the epidermis of all geckos and a few other squamates, like anoles and chameleons (Hiller 1968; Ruibal 1968; Maderson 1970; Bauer and Russell 1988; Peattie 2008; Khannoon et al. 2014). A suite of morphological specializations have evolved to control the adhesive properties of the setae and their interactions with the substrate (Russell 1979; Russell 2002; Pianka and Sweet 2005). These diverse morphologies have also informed gecko taxonomy and many gekkotan generic names refer to aspects of digital anatomy (Fitzinger 1843; Russell and Bauer 2002).

Early hypotheses of gekkotan relationships suggested the padless Eublepharidae were the sister clade to the remaining Gekkota, a scenario that implied only one or two origins of digital adhesion in geckos (Underwood 1954; Gamble et al. 2017). However, recent molecular genetic phylogenies find the Eublepharidae nested within extant Gekkota, along with numerous other padless gecko species (Han et al. 2004; Townsend et al. 2004; Gamble et al. 2011, 2015). This revised phylogenetic hypothesis
makes a single origin of adhesive digits in geckos less likely. Using near-complete generic sampling, comparative phylogenetic analyses recovered strong support for repeated gains and losses of digital adhesion where the most recent common ancestor to geckos lacked adhesive toepads (Gamble et al. 2012). However, subsequent comparative analyses, using different methods and taxonomic sampling, posit a single origin of digital adhesion in the most recent common ancestor to geckos, implying all padless gekkotans are the result of secondary loss of adhesive toepads (Hagey et al. 2017; Harrington and Reeder 2017). Distinguishing between these conflicting hypotheses, single vs. multiple origins of digital adhesion in geckos, is possible by investigating independent lines of evidence related to the evolution of digital adhesion, including: morphology, development, behavior, and genomics (Gamble et al. 2017). Detailed examination of morphology, in particular, seems to support a scenario featuring repeated gains and losses of digital adhesion (Haacke 1976; Russell 1976, 1979; Gamble et al. 2017; Russell and Gamble 2019). However, additional lines of evidence are necessary to confirm these results. Genomic studies that examine duplications of corneous beta-protein (CBP) genes may be particularly useful in resolving competing hypotheses of single vs. multiple origins of digital adhesion in geckos.

Like the rest of the reptile epidermis, setae are composed of CBPs (formerly known as beta-keratins) (Maderson 1964; Baden and Maderson 1970; Alibardi and Toni 2005; Alibardi et al. 2007; Alibardi 2016b; Holthaus et al. 2019). Most CBP genes are clustered in a single locus, the epidermal differentiation complex, that contains serially duplicated gene copies with different transcriptional orientations (Presland et al. 1989; Mischke et al. 1996; Dalla Valle et al. 2010; Greenwold and Sawyer 2010; Alfeldi et al. 2011; Alibardi 2016b). CBPs polymerize in an antiparallel orientation into linear beta-filaments which aggregate to form integumentary structures like scales, claws, and setae (Gregg and Rogers 1986; Greenwold et al. 2014). A large diversity of duplicated CBPs, each with slight sequence modifications, allows for structural diversity in the beta-filament polymers and the production of diverse integumentary structures (Greenwold and Sawyer 2011). Thus, expanding the CBP gene repertoire through gene duplication can provide the raw material for epidermal novelties, including setae (Alibardi 2009; Khan et al. 2014).

Gene duplications are an important source of evolutionary innovation (Ohno 1970; Lespinet et al. 2002). As evidence of this, duplications in corneous proteins are associated with the evolution of novel integumentary phenotypes in many amniotes (Greenwold et al. 2014; Khan et al. 2014). These include duplications in intermediate filament keratins (IF-keratins, formerly known as alpha-keratins) associated with the evolution of hair in mammals, and CBP duplications associated with feathers in birds, and scutes, the scales that cover the shell, in turtles (Greenwold and Sawyer 2011; Li et al. 2013; Khan et al. 2014). CBP gene duplications have also been associated with the evolution of digital adhesion in anoles and geckos (Alfeldi et al. 2011; Liu et al. 2015). Phylogenetic analyses of CBPs from Schlegel’s Japanese Gecko (Gekko japonicus) and Panther Gecko (Paroedura picta) genomes (Liu et al. 2015; Hará et al. 2018) show large numbers of gecko-specific duplications. However, these analyses failed to clarify whether these duplications are associated with the possible independent evolution of digital adhesion in each of these lineages, consistent with the multiple origins hypothesis, or whether these duplications arose in the most recent common ancestor of geckos, consistent with either the single origin hypothesis or, perhaps, the evolution of some other gecko-specific epidermal trait, like the ubiquitous epidermal spinules.

Thus, there are two related questions that can be addressed, in part, by examining CBP gene evolution in gecko genomes. The first, have adhesive digits evolved once or more than once in geckos? The second, has the expansion of gekkotan CBP genes coincided with the evolution of spinules or with adhesive digits and elaborate subdigital setae? These questions are interconnected and answering them requires examining the possible outcomes among the various combinations of these two questions. These outcomes can be used to compose three testable hypotheses about CBP gene evolution under these scenarios (Fig. 1). The three hypotheses in Fig. 1 illustrate simplified versions of a species tree and an associated CBP gene tree, that would support the described hypothesis. Empirical data would obviously be more complicated, as shown in the “Results” section. The three hypotheses are as follows:

Hypothesis A: Hypothesis A predicts that ancient CBP duplications are associated with the evolution of the spinulate epidermis, which is ubiquitous in geckos, not the evolution of adhesive digits and elaborate setae. Thus, both padded and padless geckos should have the full suite of duplicated CBP genes with no additional duplications associated with the evolution of digital adhesion. Under this scenario, the gekkotan epidermis is capable of developing setae through exaptation. That is, the CBPs necessary to produce setae are present in all geckos and it is only
their arrangement and number in the epidermal structures that differentiates the spinulate from the setal arrangement.

Hypothesis B: Hypothesis B predicts multiple CBP duplications associated with each independent origin of adhesive digits. Each lineage that has independently evolved digital adhesion should share a suite of duplicated CBPs, exclusive of other padded lineages that have their own set of duplications. Padless geckos should lack large numbers of duplicated CBPs.

Hypothesis C: Hypothesis C predicts ancient CBP duplications in the most recent common ancestor of geckos are associated with the evolution of setae and a single origin of adhesive digits. The CBP gene tree in this scenario is expected to look very similar to the gene tree in hypothesis A, but with an important exception. While hypothesis A predicts all geckos share the suite of duplicated CBPs, hypothesis C predicts that secondarily padless geckos should lose beta-keratin gene copies through disuse. This is analogous to the loss of IF-keratins associated in whales.
and dolphins (Sun et al. 2017) or the loss of claw CBPs in snakes (Emerling 2017).

Here, I provide a cursory examination of existing CBP data using three published gecko genomes. One species, Eublepharis macularius, lacks digital adhesion while the remaining two, G. japonicus and P. picta, have adhesive digits. Furthermore, under a scenario of repeated gains and losses of digital adhesion, G. japonicus and P. picta evolved digital adhesion independently and they have dramatically distinct adhesive toepad morphology (Gamble et al. 2012; Russell and Gamble 2019; Fig. 2). This sampling allows me to tentatively assess whether there is sufficient signal in CBP data to distinguish among the three hypotheses. However, these data are insufficient to conclusively address these hypotheses. Therefore, I also provide a framework to guide taxonomic sampling for future gecko genome sequencing sufficient to address the breadth of hypotheses describing CBP evolution described above.

**Materials and methods**

CBP sequences from Gekko gecko (Hallahan et al. 2009) and Gallus gallus (Presland et al. 1989) were downloaded from GenBank and queried against annotated CDS’s from Anolis carolinensis and G. gallus downloaded from Ensembl v95 (Frankish et al. 2017) and G. japonicus downloaded from NCBI (Clark et al. 2016). The original query sequences plus the unique BLAST results from A. carolinensis, G. japonicus, and G. gallus were used to search, against annotated CDSs from 10 genomes, including, two archosaurs: G. gallus and Alligator mississippiensis (International Chicken Genome Sequencing Consortium 2004; Green et al. 2014); one turtle: Chrysemys picta (Shaffer et al. 2013); three geckos: G. japonicus, P. picta, and E. macularius (Liu et al. 2015; Xiong et al. 2016; Hara et al. 2018); and four non-gekkotan squamates: Salvator merianae, Thamnophis sirtalis, Shinisaurus crocodilurus, and A. carolinensis (Alföldi et al. 2011; Gao et al. 2017; Perry et al. 2018; Roscito et al. 2018). The sampled gecko species represent three distinct digital morphologies: E. macularius lacks adhesive digits; G. japonicus and G. gecko have large, basal adhesive pads; and P. picta has so-called leaf-toed pads, paired pads at the distal tip of the digits (Fig. 2; Boulenger 1885). Furthermore, under a scenario of repeated evolution of digital adhesion in geckos, Gekko and Paroedura independently evolved adhesive digits, thus providing sufficient taxonomic sampling to distinguish among hypotheses A–C (Gamble et al. 2012; Russell and Gamble 2019). Given the preliminary nature of this study, I focused on using annotated genes to facilitate sequence alignment and to limit the inclusion of pseudogenes. Excluding pseudogenes is important in distinguishing between hypotheses A and C, as secondarily padless gecko species are predicted to have lost CBP genes under hypothesis C. BLAST queries, implemented in Geneious R11, used discontinuous megaBLAST, keeping a maximum of 50 hits with E-values greater than 1e–10 (Altschul et al. 1990). In cases where multiple isoforms of the same CBP gene were identified, the longest isoform was retained. Duplicate BLAST hits were removed and unique sequences aligned using a codon model in MAFFT (Katoh and Standley 2013). Gaps that occurred in >80% of sequences were removed from the alignment, and a maximum likelihood tree estimated using RAxML HPC-Blackbox 8.2.10 with rapid bootstrapping and GTR plus gamma model, implemented on the CIPRES Science Gateway (Stamatakis et al. 2008; Miller et al. 2010; Stamatakis 2014).

**Results**

The number of annotated CBP genes recovered from each species varied from six, in E. macularius, to 57, in G. gallus (Table 1). The CBP gene tree recovered an archosaur plus turtle clade and a squamate clade (Fig. 2 and Supplementary Fig. S1). Most nodes across the tree had low (<70) bootstrap support and thus, relationships should be interpreted with caution. Many species-specific clusters of duplicated CBPs recovered here were identified in previous analyses, including turtle and avian-specific duplications (Li et al. 2013; Greenwold et al. 2014). There were three sets of Paroedura-specific and Gekko-specific duplications, clades 1, 2, and 3 (Fig. 2 and Supplementary Fig. S1). The Gekko-specific duplications, in two cases, consisted of both G. gecko and G. japonicus sequences intermingled among each other (in clades 1 and 3, Fig. 1) suggesting the duplications occurred in the most recent common ancestor to the two species. The third set of Gekko-specific duplications consisted only of G. japonicus genes (in clade 2, Fig. 2), which was likely due to the limited number of G. gecko genes used here. Eublepharis macularius CBPs were scattered across the squamate clade with no species-specific clusters of gene duplications.

**Discussion**

The presence of CBP gene duplications in the two gecko lineages that likely evolved digital adhesion independently, once in the genus Gekko and a second time in Paroedura, is concordant with Hypothesis B.
Additionally, the paucity of species-specific duplications in the padless gecko, *E. macularius*, is also consistent with either hypotheses B or C. *Eublepharis macularius*, had approximately the same number of annotated CBPs as the snake, *T. sirtalis*, and the lizards *S. merianae* and *S. crocodilurus*. Differences in the relative number of annotated CBP genes among bird and reptile species largely matches previous findings (Alföldi et al. 2011; Li et al. 2013; Liu et al. 2015; Hara et al. 2018). Species with novel integumentary phenotypes, e.g., feathers, shells, and adhesive digits, having more duplicated genes than species lacking such novelties. However, these numbers should be interpreted cautiously as they were derived from annotated CBPs genes and not all of which are likely to be annotated in these genome assemblies. The number of annotated CBPs recovered from several species differed from previously published accounts. For example, whole genome scans recovered 71 *G. japonicus*, 23 *A. carolinensis*, and 120 *P. picta* CBP genes (Alföldi et al. 2011; Liu et al. 2015; Hara et al. 2018). We found only 26 *G. japonicus*, 19 *A. carolinensis*, and 53 *P. picta* CBPs among the annotated genes. The presence of numerous unannotated CBP genes was alluded to in Hara et al. (2018) as they could only find 32 *G. japonicus* and 14 *A. carolinensis* CBPs for their phylogenetic analysis. Thus, relying solely on annotated genes will underestimate the total number of beta-keratins in the genome. It is also possible that the quality of a particular genome assembly

Fig. 2 (A) Maximum likelihood phylogeny of annotated CBP genes from 11 reptile and bird species. Shaded circles at the tips indicate species assignment. Circles at internal nodes indicated bootstrap support, black circles have values >70 while white circles indicate values <70. Scale bar units are substitutions/site. Clades 1, 2, and 3 are discussed in the text. An alternate version of this tree, that includes taxon names and gene ID’s, is included in the Supplementary Materials (Supplementary Fig. S1). (B) Foot of *G. gecko*, which has a similar morphology with *G. japonicus*. The black bar indicates the single, subdigital adhesive pad that extends along most of the digit length, a so-called basal pad. (C) Foot of *P. picta*. The black bar indicates the location of the paired adhesive pads at the digit’s distal tip, a so-called leaf-toed pad. (D) Foot of *E. macularius*. Note the lack of adhesive pads.
could negatively affect gene discovery. However, I found no relationship between the number of annotated CBP genes recovered and two estimates of genome assembly quality: the number of annotated protein-coding genes; and scaffold N50 (Supplementary Fig. S1). Future efforts should focus on BLAST searches of the entire genome with subsequent screening for open-reading frames to find all functional gene copies.

While the hypotheses presented here appear distinct in their simplified form (Fig. 1), aspects of hypotheses A, B, and C are not mutually exclusive. For example, there are several padless gecko species that clearly exhibit a secondary loss of adhesive digits (Lamb and Bauer 2006; Gamble et al. 2012). Loss of CBPs in these secondarily padless species, consistent with aspects of hypothesis C, might be expected even if the broader dataset supports another hypothesis, hypothesis B, for example. Similarly, there may be some early CBP duplications associated with the evolution of spinules in geckos, hypothesis A, even if most remaining duplications are associated with lineages that have independently evolved digital adhesion, hypothesis B. Thus, teasing apart various aspects of these hypotheses might involve additional analyses, beyond just an examination of gene trees. Utilization of model-based analyses that count gene gains and losses in a phylogenetic context would be extremely useful in this regard. A variety of methods can infer gene duplication and loss by reconciling a gene tree with a species tree (Vernot et al. 2008; David and Alm 2011;
Boussau et al. 2013; Górecki and Eulenstein 2014) or model gene copy gains and losses with a birth and death stochastic process (Hahn et al. 2005; Librado et al. 2012; Han et al. 2013). Additionally, examining the spatial orientation of these genes on assembled genome scaffolds can help clarify syntenic and gene orthology among CBP genes. Finally, investigating where specific CBPs are expressed in chicken, Anolis, G. gecko, and other species has proven especially useful in understanding the function of many duplicated beta-keratin gene copies (Hallahan et al. 2009; Dalla Valle et al. 2010; Ng et al. 2014; Strasser et al. 2014). Determining where different CBP genes are expressed, using comparative RNAseq and in situ hybridization, e.g., Alibardi (2013, 2016a, 2018) and Alibardi et al. (2007), will aid in understanding the functional role of each gene and further clarify the relationship between CBPs and digital adhesion.

While pseudogenes should be eliminated when estimating the number of gains and losses of functional gene copies associated with the evolution of a particular trait, pseudogenes, when present, are still informative. Several methods exist to estimate the timing of gene inactivation, which can help determine when a trait associated with that gene was lost (Emerling and Springer 2014, 2015). Such analyses would prove useful in testing hypotheses about beta-keratin loss associated with a secondary loss of digital adhesion, such as hypothesis C. Similarly, such techniques could be used with secondarily padless species nested within a clade of species with adhesive digits, e.g., Lucasium dameum, Pachydactylus rangei, or Chondrodactylus anguifer (Lamb and Bauer 2006; Gamble et al. 2012).

Rigorously testing the above-mentioned hypotheses using whole-genome data requires adequate taxonomic sampling. At a minimum, this means sampling species both with and without adhesive digits from clades that span extant gekkotan diversity. Sampling multiple lineages that are hypothesized to have independently evolved digital adhesion is necessary to adequately distinguish hypothesis B from hypothesis C. These lineages can be identified by examining the phylogenetic ancestral state reconstructions from Gamble et al. (2012) and Russell and Gamble (2019). While this sampling strategy is sufficient to distinguish among hypotheses A, B, and C, the inclusion of several additional species could provide deeper insight into the relationship between digital adhesion and the gain and loss of CBP genes. For example, sampling one of several species that have recently evolved digital adhesion might offer insight into the earliest stages of lineages-specific CBP duplications. *Gonatodes humeralis*, for example, has been shown to have recently evolved digital adhesion and is nested within a clade of species that lack adhesion (Russell et al. 2015; Higham et al. 2017). Even under a scenario where digital adhesion evolved just once, in the most recent common ancestor of geckos, *G. humeralis* would almost certainly represent an independent gain of digital adhesion following a loss in the most recent common ancestor to *Gonatodes*. Thus, sequencing the genome of *G. humeralis* and any other *Gonatodes* species would provide an outstanding comparison between a padless gecko and a recently evolved digital adhesive mechanism. Similarly, sequencing the genome of species that has recently lost digital adhesion would provide an important test of whether loss of adhesion is accompanied by the concomitant loss of CBP genes. To address this question I include *P. rangei*, a gecko that has secondarily lost adhesive digits (Lamb and Bauer 2006), along with the closely related *Pachydactylus geitje*, that retains digital adhesion. A potential sampling strategy that incorporates all of these suggestions is illustrated in Fig. 3. This is intended to be a just one example of many possible sampling plans that could address these questions.

Phylogenetic comparative methods are powerful tools to generate testable evolutionary hypotheses (Pagel 1999; Nunn 2011). However, they are not without their flaws and conflicting hypotheses may find support using different methods and data (Schluter et al. 1997; Maddison and FitzJohn 2015; Rabosky and Goldberg 2015). In the case of digital adhesion in geckos, comparative methods have

| Table 1 The number of annotated CBP genes identified from the BLAST analysis of 11 listed species |
|-----------------------------|-----------------------------|-----------------------------|
| Species                     | Number of CBP genes | Source                     |
| Alligator mississippiensis  | 11                  | NCBI                       |
| Gallus gallus               | 57                  | Ensembl                    |
| Chrysemys picta             | 25                  | Ensembl                    |
| Anolis carolinensis         | 19                  | Ensembl                    |
| Gekko gecko                 | 20                  | NCBI                       |
| Pareodura picta             | 53                  | transcriptome.cdb.riken.jp/reptiliomix/ |
| Eublepharis macularius      | 6                   | http://ggadb.org/dataset/100246 |
| Gekko japonicas             | 26                  | NCBI                       |
| Shinisaurus crocodilurus    | 7                   | http://ggadb.org/dataset/100315 |
| Thamnophis sirtalis         | 7                   | NCBI                       |
| Salvator merianae           | 8                   | http://ggadb.org/dataset/100529 |

These genes were used to generate the phylogeny in Fig. 2. The source of annotated genes is listed in the final column.
supported two conflicting hypotheses: a single vs. multiple origins of digital adhesion. To resolve this conflict, other sources of data are needed. Preliminary data support a hypothesis where CBP gene duplications in geckos are associated with the repeated origins of digital adhesion (Hypothesis B). However, further genome sequencing is necessary to robustly test competing hypotheses. Additional gecko genomes are sure to be sequenced over the next few years and these new data will be invaluable for settling this ongoing debate.

**Acknowledgments**

Thanks to T. Higham, A. Stark, and T. R. Russell for organizing the SICB symposium and editing this special issue. Thanks to A. Griffing, C. Richards-Zawacki, and two anonymous reviewers for helpful comments and S. V. Nielsen for photos pf *Phylloclactus wirshingi* and *G. gecko*.

**Funding**

Support for this research is from the U.S. National Science Foundation (NSF-DEB1657662).

**Supplementary data**

Supplementary data available at ICB online.

**References**


Duplications in corneous beta protein genes


