A Numerical Analysis of Isoelectric Focused Keratin Monomers: Affinities of Some Western Indian Ocean Green Geckos

R. S. THORPE
Department of Zoology, University of Aberdeen, Aberdeen AB9 2TN, U.K.

Key Word Index—Phelsuma; geckos; keratin; isoelectric focusing; OTUs.

Abstract—This is a case study in the numerical analysis of intraspecific and interspecific differences in isoelectric focused proteins. The proteins are epidermal keratin monomers extracted and characterized by relatively novel procedures. The numerical analysis of the biochemical data by principal component analysis indicates the relative similarity of the populations whilst Wagner trees were computed to hypothesize their phylogeny. The character states of the tree nodes ('ancestors') are deduced and the phylogenetic tree is fitted directly onto the ordination diagram. In this way evolutionary divergence, convergence and extent of change can be readily visualized. Little convergence exists in this biochemical data set, there being close agreement between the phenetic and phylogenetic analyses. The existence of three species is indicated and this contradicts some facets of the conventional classification.

Introduction
This paper is intended as a case study in the numerical analysis of biochemical data from isoelectric focusing. Whilst the numerical analysis of allozyme frequencies to investigate intraspecific evolution is well documented, the numerical analysis of isoelectric focused proteins to investigate interspecific affinities is less well developed [1].

Isoelectric focused (IEF) proteins often show interspecific differences, e.g. salmonid fish [1] but photographs of the gels are generally presented just for visual comparison and no quantification is given. Alternatively a specific band may be pinpointed in order to demonstrate a difference in IEF profiles of a pair of samples as in Jamieson and Turner's [2] study of two species of eel. These procedures are satisfactory for identification but do not allow a quantified comparison of several species or races so that their evolutionary relationships can be hypothesized and/or a taxonomy derived.

This study quantitatively compares IEF profiles of five OTUs (operational taxonomic units) so that their phenetic relationships (relative similarity) can be portrayed and also their evolutionary relationships hypothesized. The IEF profiles of the hypothetical 'ancestors' can be deduced and the evolutionary tree and phenetic relationships are combined on a single diagram so that any evolutionary divergence and convergence can be easily visualized.

The biochemical technique employed was relatively novel. The biochemical affinities were based on isoelectric focused keratin monomers derived from the shed epidermis of geckos. Previous studies [3] have shown the value of this biochemical technique for investigating the systematics of reptiles. These keratin monomers may, at least in birds, represent single gene products [4]. Consequently, the technique may be considered as giving an estimate of 'genetic affinity' but it should be noted that the systematic value of the technique does not depend on this assumption.

The organisms studied are green, or day, geckos (Phelsuma). They are diurnal, arboreal, omnivorous lizards found in Madagascar and on many Indian Ocean islands. There are over twenty nominal species in this genus but their validity, and interrelationships are generally unknown or subject to disagreement.

This study aims to investigate the relationships of five OTUs from the north west of the range. These five OTUs may belong to the 'madagasariensis' groups of species [5] but the composition and existence of such a species group is not sure. Depending on the author, these five

The OTUs are: (a) *P. madagascariensis* from Madagascar, the specific status of which is uncontroversial. References to the 'Madagascan population' pertain to this species only and not to other Madagascan species. (b) A population from Mahe, the largest of the Granitic Seychelles. This population is sympatric with, and distinct from, *P. astriata* [7]. It has variously been referred to as *P. pulchra* [8], *P. madagascariensis longinsulae* [9], *P. abbotti pulchra* [6] or *P. longinsulae pulchra* [5]. (c) A population from Praslin, the second largest of the granitic Seychelles. This population is sympatric with, and distinct from, *P. astriata* [7]. It has been described as *P. sundbergi* [8] or *P. madagascariensis sundbergi*. Any reference to the populations of the granitic Seychelles (Praslin/Mahe) does not pertain to *P. astriata*. (d) A population from Aldabra. There is only one species of *Phelsuma* on this coralline island in the outer Seychelles. It has been referred to as *P. abbotti* (abbotti) [8] or *P. madagascariensis abbotti* [9]. (e) A population from Assumption. There is only one species of *Phelsuma* on this coralline island in the outer Seychelles. It has recently been described [10] as *P. abbotti sumpto*.

These five taxa represent three geographic regions: (i) Madagascar, (ii) the granitic Seychelles (Mahe and Praslin), and (iii) the outer coralline islands of Aldabra and Assumption. The geographic relationships are seen in Fig. 1.

The previously described species frequently bear little relation to the geographic position of the populations. The populations from the geographically close islands of the granitic Seychelles (e.g. Mahe and Praslin) have never previously been thought to be a single species distinct from other populations considered here. Even recent studies [5, 10] consider them to be separate species. However, they have often been linked with geographically distant populations. For example, Mertens [6] has given the Praslin population conspecific status with the distant Madagascan population whilst considering the Mahe population to be conspecific with the distant Aldabran population. There are a large number of different nomenclatures for these taxa; they are reviewed in Gardner [11].

The level of sympatry of island *Phelsuma* is low and consequently the biological species concept is of limited help in validating species. However, two of the five OTUs are nominally sympatric. *P. abbotti* from Aldabra has a nominal conspecific on Madagascar which is sympatric with *P. madagascariensis*.

**Results**

Comparison of the figured gel (Fig. 2) with other gels allowed over forty isoelectric focused band positions to be recognised. Seventeen of these positions (identified as A–Q) show variation in the presence/absence of a band between taxa. The band pattern is diagrammatically represented in Fig. 3 and the resultant 17 character × 5 OTU matrix is given in Table 1.

The relative similarity of the isoelectric focused SCMK patterns, as portrayed by principal component analysis (Fig. 4), shows three distinct phenenetic categories: (a) The Madagascan population. This population is distinct from the other four OTUs. (b) The granitic Seychelles group. The population of Praslin and Mahe are very similar to one another but distinct from the other three OTUs. (c) The Aldabra/Assumption group.
FIG. 2. PHOTOGRAPHS OF ISOELECTRIC FOCUSED GELS. MH = Mahe, PR = Praslin, MD = Madagascar, AL = Aldabra, AS = Assumption.
A to Q indicate band positions. X1, X2 and X3 are heavy bands. See text for explanation.
These two populations are very similar to one another but distinct from the other three OTUs. The Wagner tree shown in Fig. 4 has three lineages. These three lineages coincide with the three phenetic groups, there being little parallelism or convergence. The mid-point root of the tree is the node common to all three lineages.

One other, equally parsimonious, tree is possible. This tree is only trivially different from the figured tree. The mid-point root is one unit away from that of the figured tree (i.e. one unit from the node joining the Madagascan and Assumption/Aldabran lineages rather than directly at that node) and the position of the Assumption and Aldabra populations within their lineage is reversed. However, the basic picture of three diverging lineages remains.

The extent of change along the granitic Seychelles lineage and the Madagascan lineage is equal but there is less divergence along the Assumption/Aldabra lineage. A fuller analysis, using more species, is necessary to confirm this disparity in extent of evolutionary change.

According to the figured Wagner tree the seven changes in the 'Madagascan' lineage involve gaining a band at positions E, K and N and losing a band at positions B, H, L and Q. The

---

TABLE 1.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Madagascar</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Assumption</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Aldabra</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Mahe</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Praslin</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
five changes in the granitic Seychelles lineage (going from the root to the common Mahe/Praslin node) involve gaining a band at positions D and I and losing a band at positions F, J and P. The two changes from the Praslin/Mahe node to the Mahe population are the gain of a band at C and the loss of a band at G whilst the single change to the Praslin population is the gain of a band at O. The two changes along the Assumption/Aldabra lineage are the loss of a band at A and the gain of a band at M. The single change from Assumption to Aldabra population is the loss of a band at Q. These changes related specifically to the figured Wagner tree and will differ in the alternative tree. In particular, the direction of evolution between Assumption and Aldabra differs in the two schemes and is consequently unresolved.

Investigation of the changes at individual band positions reveals no substantial convergent evolution. Only one band changes convergently in two separate lineages, i.e. a band is lost at position Q in both the Madagascan lineage and in the lineage between Assumption and Aldabra.

In the alternative tree there is still only one case of convergence. In this tree a band is convergently gained at position Q in both the granitic Seychelles lineage and the lineage from Aldabra to Assumption.

Discussion
The lack of an outgroup species means that it is not possible to rigorously define the five OTUs as a member of monophyletic 'madagascariensis' species group. Nevertheless a comparison between the gels presented here and those of other species presented elsewhere [3, 12] show that the five OTUs have some distinctive features in common which some other species lack. For example, three heavy bands at circa pH 5, labelled X1, X2 and X3 can be seen in all five OTUs both in Fig. 2 and in Thorpe and Giddings (Fig. 3 samples C, D, E i.e. Praslin, Mahe and Aldabra). These are usually lacking in species not considered to be the 'madagascariensis' group, i.e. samples B, F and G in Fig. 3 of Thorpe and Giddings [12]. However, P. atriata [12] (Fig. 3, sample H) also appears, at least superficially, to have rather similar heavy bands although it is not considered to be a member of the 'madagascariensis' group. Detailed biochemical work on the Praslin sample indicates that these three heavy bands represent isoelectrically coincident high sulphur and low sulphur keratins [12] (Fig. 2, samples C and D). Consequently, it is an oversimplification to consider these bands as representing three genes which define a 'madagascariensis' species group even though the keratins may be useful in clarifying sub-generic relationships.

However, the relationships between the five OTUs seem unequivocal. They clearly fall into three phenetic groups which coincide with the three clades. It is the almost total absence of convergence in this biochemical data set that results in this close agreement.

These three groups closely reflect the geographical relationships of the samples in a way previous studies fail to. The Mahe and Praslin populations are both from the central granitic Seychelles and are shown by this study to be very closely related. No previous study [5, 6, 8-10, 15] has noted this close relative similarity. This is probably due to the substantial difference in body size of the lizards from these two islands. Also Assumption and Aldabra are geographically close coraline islands and their populations are, biochemically, very similar to one another.

Insofar as the formal taxonomy is concerned the Madagascan sample is uncontroversially recognised as P. madagascariensis. Cheke [10] recognises the Aldabran and Assumption populations as conspecific (P. abbotti) with one another but separate species to the other three taxa. The biochemical evidence is in agreement with this.

Since P. abbotti and P. madagascariensis are sympatric in Madagascar this enables one to fix a minimum rank of 'species' for these two phenetic groups/lineages. Since the granitic Seychelles lineage/phenetic group (i.e. Mahe/Praslin) is of equal status to these other two taxa (i.e. P. madagascariensis and P. abbotti) one can also recognise it as being at least the rank of 'species'. As for the maximum taxonomic rank, collecting trips to numerous Seychelles islands by Gardner and myself and the subsequent study by Gardner [pers. comm.] shows that the forms are conspecific. Taking these two points together the granitic Seychelles taxon can be seen to be at the level of species.

The five OTUs therefore fall into three taxa irrespective of whether one uses phenetic or
cladistic criteria and these three taxa have the rank of species. The Mahe and Praslin populations have not previously been recognised as being conspecific with one another and distinct from other Madagascan or Aldabran species.

Given the rules of nomenclature the specific names for the granitic Seychelles species (i.e. Praslin/Mahe) could be either *P. sundbergi* (type locality, Praslin) or *P. longinsulae* (type locality, Long Island, which is a small islet off Mahe). Gardner [pers. comm.] has chosen the specific name *P. sundbergi* because the type locality of ‘*longinsulae*’ is now an inaccessible prison island.

**Experimental**

*Biochemical methods. The keratin of reptilian epidermis is an extremely resistant protein polymer but it may be broken down into soluble monomers with the appropriate biochemical techniques. The *S*-carboxymethylated keratin (SCMK) monomers are obtained by taking shed epidermis, washing it, solubilizing it in a mixture containing thioglycollic acid, carboxymethylating with iodoacetic acid, and then dialysing and concentrating. The SCMK monomers are highly heterogeneous and can be characterized by isoelectric focusing on specially prepared gels. The rather involved biochemical procedure is described in detail by Thorpe and Giddings [3].

Chemical fractionation and two-dimensional electrophoresis [3, 12] have indicated some isoelectric coincidence in the single-dimension isoelectric focusing. Nevertheless a large number of monomers can be characterized by their isoelectric point. Depending on the quality of the gel and its pH range the number of bands can be over forty. The technique was originally developed with the intention of using epidermis from museum specimens but in this study fresh shed epidermis was used.

Epidermis from different individuals was not pooled so each sample represents an individual specimen. Analysis of a range of specimens from several populations revealed polymorphism in only one band in one population [3].

*Encoding IEF patterns. The isoelectric profiles can be compared by the presence or absence of a band at a given pH (position), there being no need to resort to comparing band concentration. Because of the fineness and proximity of many of the bands it is best to compare the original gels rather than a photograph and permute the positions of the samples so that all pairs of samples are adjacent at least once. Recording the presence or absence of homologous bands is difficult and one is required to be painstaking.

The presence of a band at a given pH is coded as 1, absence as coded as 0 and bands that are present in all five taxa are ignored. One consequently obtains an OTU x character matrix of binary data in which all characters are variable. No further data transformation was carried out.

*Numerical analysis. The relative similarity of the taxa was portrayed by principal component analysis. A between-character covariance matrix is computed and the eigen-vectors extracted from this r mode matrix. The resultant component scores give an identical ordination to a principal coordinate analysis on a taxonomic distance (or mismatch distance coefficient) matrix derived from untransformed (binary) data [13].

The hypothesized phylogenetic relationships were derived from a minimum length Wagner tree [14] from the same character x OTU matrix used for the principal component analysis. As the number of taxa is low it was possible to carry out an exhaustive search for minimum length trees using a modified version of Colless’s ‘Wagnets’ and ‘Pendek’ programs. The character states at the HTUs (nodes) were hypothesized and the tree marked to indicate the gain or loss of a band.

Given the present conditions the Wagner tree can be fitted directly onto the principal component ordination diagram so that the phenetic and cladistic relationships are comparable. The Component score of a given node (HTU) is found as the sum product of the character state vector and the appropriate eigenvector.

It was not possible to outgroup root this tree because of the combination of two situations. First, the relationships within the genus are not well enough known to select a single outgroup species and second, it would have involved a disproportionately large amount of work to use many outgroup species because of the need to have each pair of samples adjacent on the isoelectric focusing gel.

Acknowledgement – I would like to thank Professor F. Robertson (Genetics, Aberdeen) for reading the manuscript.

**References**