Controversial Timing for the Arrival in Australia of the Boab (*Adansonia gregorii*).

JD Pettigrew FRS

Emeritus Prof. JD Pettigrew

j.pettigrew@uq.edu.au

6842 words
Abstract: The Australian baobab, or Boab, (*Adansonia gregorii*), is mysteriously separated from the remaining 8 African and Madagascan species in its genus by the Indian Ocean. Information about the time of the Boab’s arrival in Australia was sought using the molecular genetics of the nITS gene in 220 Boabs sampled from their complete distribution across the Kimberley. nITS substitutions had limited diversity (0.36%), in keeping with the “young” genetics shown by DNA microsatellites in another study. There were more substitutions in NW trees compared with those at the Eastern limit of its distribution around the Victoria River, which are presumed to be more recent. Coalescence time and divergence time could be based on reliable data concerning the range of nITS substitutions in the Boab and their divergence from the African baobab sister species, but these estimations had a stumbling block: there is no reliable estimate of the mutation rate for the nITS gene of the Boab. Earlier publications made the assumption that the mutation rate for the Boab was the same as that of its African sister taxon (*Adansonia digitata*). This risky assumption gives a divergence time of 2.4 Ma that is discrepant with other studies, such as microsatellites, that indicate a much younger age. This assumption also overlooks the completely different ecology and pollination of the Boab compared to its African sisters, as well as the known lack of any phylogenetic correlation for nITS mutation rate. In contrast to the assumption of equality between African and Australian mutation rates, three lines of qualitative and quantitative evidence refute this assumption and indicate that the Boab is an outlier with a very high mutation rate. Time dependence of mutation rate predicts an elevated mutation rate as does Yang’s relative rate test of mutation rate. A quantitative phylogenetic study of nITS in all 9 baobab species, with an appropriate outgroup, shows that all Boab samples have branch lengths that are 18X longer than any other baobab species in the phylogeny. Combined with the base mutation rate of 2.5 x 10^-9 substitutions/site/year, this gives a new estimate of the Boab mutation rate of 45 x 10^-9 substitutions/site/year.

The new estimate of mutation rate gives a coalescence time and a divergence time of around 72 Ka, in line with the Toba mega-eruption that marked the beginning of modern human migrations out of Africa, one of which may have brought a Boab progenitor to Australia.
Fig. 2: The Boab, *Adansonia gregorii*. Flowers (a), Buds (b), Seed pods (c), Characteristic baobab pachycauly (d,e).

Introduction:
The Australian baobab, or Boab, *Adansonia gregorii*, is a genuine botanical mystery. The puzzle might well be better understood if we knew more about when it arrived in Australia. The first use of molecular genetics to examine its timing, effectively dismissed the idea that the Australian baobab had separated from African or Madagascan relatives as a result of continental drift. The estimated Australia-Africa divergence from molecular genetics was 2.4 million years compared to the more than one hundred million years expected from vicariance (Baum et al. 1998). This result from molecular genetics showed the effectiveness of molecular phylogeny in settling some issues of timing, but questions remained. First, only a single sequence was available from the Boab, so there were not enough data available to decide whether its origin was amongst the Madagascan, or the African, species of baobabs. At the time, Madagascar was a popular choice of the two, because the Boab is a diploid like all six Madagascan baobab species and unlike the African tetraploid (Baum and Oginuma 1994). The second reason favoring a Madagascan origin was that a flotation route from Madagascar to Australia had been established by *Aepiornis* (elephant bird) eggs that had made the journey on two recorded occasions (Long et al. 1998). Such phylogenetic uncertainty about African vs Madagascan origins makes it very difficult to apply the relative rate test, as shown below. A second major uncertainty concerns the mutation rate to be applied to the divergence. In the study by Baum et al. (1998) and in subsequent studies of Boab timing (Bell et al 2014, Rangan et al. 2015), it was assumed that the Boab’s Australian mutation rate, is the same as the African baobab, *Adansonia digitata*. No basis for this assumption has been advanced, a possible weakness of which can be shown by a simple exercise in the time-dependence of mutation rate (see below). This significantly further shortens the apparent calibration time of 2.4 Ma suggested by the assumption of equality between Australian and African mutation rates. Application of a correction from the time-dependence curves of Ho et al. (2005) applies to all genomes although it is less well known for plants (Drake et al. 1998). An estimate of the correction would be in the range of 0.5-1.5 Ma, giving a divergence time of 0.9-1.9 Ma instead of 2.4 Ma. The exact magnitude of this correction is not as important as the question it raises about the assumption of equality of Australian mutation rates with those of the putative African baobab precursors. This correction yields a younger age for the Boab, as do other approaches such as DNA microsatellites, raising the possibility that the 2.4 Ma divergence is an overestimate based on the assumption of a low rate.
This study attempts to go beyond the question of the Boab mutation rate to provide a more precise value. At the same time, more evidence is presented for nITS substitutions in a large number of Boabs. Two qualitative techniques imply an elevated mutation rate for the Boab, and a third quantitative technique shows a 18X elevation in mutation rate that gives an arrival time of around 72 ka, consistent with circumstantial evidence of human transport.

Methods and Materials:

Choosing the Mutation Rate:

The compilation of nITS data on substitutions in a large sample of individual trees is an important prerequisite to the estimation of their coalescence or divergence time. Even more important, however, is the mutation rate so that the number of substitutions in the phylogenetic reconstruction can be converted to time. There is a widely-used mutation rate for the African baobab, *Adansonia digitata*, which is an average of two rates, for cotton, genus *Gossypium* (Wendel et al 1955), and for the family Winteraceae (Suh et al 1993). These two rates are separated by a factor of ten, but were chosen because each had a well-documented phylogenetic history, that stretched back to the early history of angiosperms in the case of Winteraceae (Suh et al 1993). The choices were made largely on the basis of phylogeny because baobabs lack fossilised wood (it is too friable) and a clear cut fossil pollen record, so the published mutation rate (Baum et al 1998) is a compromise between the rate from ancient Winteraceae (0.45/site/year) (Suh et al 1993) and the rate derived from cotton (45/site/year) (Wendel et al 1955). The ten-fold difference reflects the difficulties and imprecision of mutation rate estimates in the context of the present study where distinguishing between the different scenarios for the arrival of the Boab would require precision in the mutation rate estimate. The range would be even greater if other angiosperms with well-documented phylogeny and mutation rate are included in the estimate of rate, such as Primulaceae (8.4) and Gentianaceae (19), which would increase the range of possible rates in the African baobab to 40x. This context of wide variations possible in nITS mutation rates needs to be borne in mind in assessing mutation rate of the Boab, *ab initio*.
The estimate for mutation rate of *A. digitata* hinged upon phylogenetic information from from
distant, ancestral lineages. In the case of the mutation rate for *A. gregorii*, information about
phylogeny can now be obtained from much closer to the stem arising near those of its sister taxa,
as uncertainties have now been resolved, such as the identity of the crown taxon (which is *A.
digitata*, not one of the Madagascan taxa) as well as the vexed question whether the Australian
baobab is closer to the African or to the Madagascan clade (recent evidence indicates that it is
the African).

These are crucial, but fragile, details that are essential for the following phylogenetic analysis
because of evidence that they are altered by details of the reconstruction, such as an
inappropriate choice of outgroup or variation in the treatment of indels (Pettigrew et al. 2012).
The phylogeny of all 9 baobabs has already been published, and some criticism has been leveled
at its use here. The point is that this is the first time that the baobab phylogeny has been used to
estimate mutation rate. The late recognition of the importance of new analysis is the result of the
easy and unquestioned assumption that has been made repeatedly in the literature, which has
assumed that Boab’s nITS mutation rate is the same as the African baoab. The author was also
seduced by this assumption at first:- hence my co-authorship of the papers that use this

The key observations are completely new interpretations on the branch lengths of each baobab
species, never mind the fact that the overall phylogeny has been published. These show that most
baobabs in the phylogeny have similar branch lengths, in partial support of the assumption that
different baobabs could be assumed to have the same mutation rate. However, this was not true
for the Boab, which as an obvious outlier, with branch lengths that were 18x longer than any of
the other 8 baobabs in the phylogeny. The implications of this much greater rate of change in the
Boab are that its mutation rate is seriously underestimated by assuming that it is the same as the
African species, *A. digitata* and that arrival time may have been 18x earlier than the 2.4 Ma
estimate based on that assumption. Arrival may have been as early as 1.4 x 10^{-6}/18 = 78 Ka.
gregorii clearly has a greater rate in both of these diagrams, where length is proportional to divergence.

Calculating the exact rate increase for *A. gregorii* requires a judgement about the exact phylogenetic position of the node (n), where *A. gregorii* leaves the stem. In detailed phylogenetic assessments of the phylogenetic position of *A. gregorii*, there has been uncertainty about the degree of closeness of the relationship between *A. gregorii* and the African baobabs, but there is now a consensus that *A. gregorii* is closer to the African baobabs than to the Madagascan baobabs. The calculated increase in rate could therefore vary from 2X to 18X as different analyses place this node nearer or closer to the African clade, with the highest value of 18X when the node is closer to the African clade.

Relative Rate Calculation:

The classical test for rate variation is attributable to Yang (2002). A graphical demonstration of Yang's approach to the present problem is shown in Fig. 1. The figure makes it obvious that the Boab, *A. gregorii*, has an elevated rate. Estimating the magnitude of that elevation would be necessary for any attempt to determine timing using the pattern of substitutions in nITS. Unfortunately, such precision is limited by an aspect of the phylogenetic information: viz:- its uncertainty.

The relative rate test requires knowledge about the phylogeny, particularly the outgroup against which most measurements are made.

Phylogenetic reconstructions of of baobabs are fragile because only a small number of informative substitutions are available. As an example of this fragility, there are two published versions of the phylogeny that could not be further apart from each other. In one phylogeny, the only tetraploid baobab is found at the base of the phylogenetic tree, with the other seven species therefore having had to undergo multiple, independent reversions to the diploid state (Baum et al. 1998). The alternative phylogeny, which is generated when a different outgroup is chosen, or the numbers of replicate samples from different baobabs are balanced, has the tetraploid at the crown, along with the Australian and African species (Pettigrew et al. 2012). This stark difference can
be explained by the forcing effect of the outgroup, but for the present analysis the important
point is that relative rate is not an absolute measure, but is affected by the phylogenetic
relationships of the taxa used for the relative rate test. This is shown by the comparison in Fig. 1,
where the relative rate of *A. gregorii* is much greater in comparison to the African species (A)
when its node is close to them, than it is when the node is close to the Madagascan clade (B).

It would be possible to make an arbitrary choice between these possibilities and use that to
calculate a relative rate. Such a dichotomous choice would inevitably lose information, so we
chose an alternative that reflected the posterior probability of the phylogeny (using Mr. Bayes)
and compared that to the raw rates obtained using PAUP. Although this phylogeny has already
been published (Pettigrew et al 2012) this is the first time that there has been a detailed
examination of the branch lengths of each of the many baobabs on the phylogeny. This is shown
in Fig. 2 which allows a quantitative comparison between rate increases and the background. The
average relative length of *A. gregorii* branches in PAUP was 18x the baseline values obtained
using Mr. Bayes. Since the base line calculation assumed a rate of 2.5 x 10^{-9}, by this
analysis, the new estimate for the mutation rate of *A. gregorii* is 4.5 x 10^{-8}
(changes/nucleotide/year).

Time Dependence:
The revised Australian mutation rate fixes a coalescence time in Australia of 72 ka. This short
time is well inside the period of exponentially increasing mutation rate that has been defined in
the curves of Ho and collaborators (2005) of time dependence of mammalian mtDNA mutation
rate. Time dependence is a phenomenon that affects all genomes, even viral genomes that have
revealed details of the underlying mechanism, where there is a delay between "purifying
selection" and the preceeding initial mutations which are therefore more numerous at short time
intervals. Although these curves from mammalian mtDNA do not have calibrated scales,
quantitation is possible using data from plant taxa that have both high mutation rate and very
short calibration times, such as *Gentiana var. Ciminalis*, which has the highest published nITS
rate, although somewhat less than the rate we estimated for *A. gregorii* (Kay et al. 2006).
The exercise has two results:-
1. Although our estimated rate for *A. gregorii* is high, based on phylogenetic reconstruction, it
corresponds to the values given by the quantitative time dependence curves from mammalian mtDNA (Ho et al. 1988). No work on plants has provided details of time dependence of comparable precision, but the phenomenon, originally discovered by Crow, is found in all genomes, from prokaryotes to eucaryotes, and in nuclear genomes as well as mitochondrial genomes, so it can be expected that this large increase in mutation rate is valid, even if the quantitative details in Boabs are not known with precision (Drake et al. 1998).

2. This result establishes the plausibility of such high rates, near the upper end of the known range of plant rates (3-50 x 10^-9/site/year) (Wolfe et al. 1987).

Phylogenetic reconstruction:

Since the relative rate calculation was dependent on the underlying phylogeny, a number of different reconstructions were used try to minimise the interaction between relative rate measurements and the particular phylogenetic reconstruction. Previous work had shown that phylogenetic reconstructions of baobabs are fragile, based upon a small number of informative substitutions (Pettigrew et al. 2012). For this reason, the topology of the phylogeny could be radically changed by varying the outgroup or failing to balance the numbers of taxa. For example, Madagascan baobabs appear at the crown of the phylogeny rather than its base if one uses more than two replicates of the Madagascan sequences. Such an “inversion” of the phylogeny, where the tetraploid baobab appears at the base, where many unlikely reversions to the diploid state would be required, also occurs with some inappropriate outgroups and is a feature of some published phylogenies (e.g. Baum et al. 1998). This raised a question as to the appropriate reconstruction on which to base relative rate calculations, since some alternative phylogenies, such as the “inverted” topology, reduce the magnitude of any rate increase.

DNA Collection and Extraction: Emergent leaves were sampled from 220 trees over the complete present distribution of A. gregorii in the Kimberley, which has notable gaps, such as the central plateau (where Boabs are eliminated by occasional frost), and the western coast of Admiralty Gulf (where uninterrupted stone slabs preclude Boab growth) (Fig. 2b). Accessible trees near roads and tracks were sometimes sampled from the roof-rack of a 4WD using a 4 m fruit picker, but most samples were taken in flight from a hovering R22 (Robinson) helicopter. The cut leaves were placed in a zip-lock plastic bag with 15 g silica gel. DNA extraction was
based upon the Qiagen DNeasy Plant protocol with a few modifications: viz: 1. Leaves were
ground (without need for liquid N2 because of “glassy desiccation” that affects baobab leaves in
silica gel), with an equal mass of PVP (polyvinylpyrrolidone); 2. After the precipitation step,
tubes were spun at 20,000xg for 30 min at 4 °C to give a clearer separation of the supernatant. 3.
DNA was eluted from the spin columns three times, each with 100 µl of buffer AE. Primers
ITS.LEU (gtcactgaaccttacatttag) and ITS4 (tcctcgettattgatagc) were used to amplify and
sequence the nuclear ribosomal internal transcribed spacer region (nrITS). Primers for
chloroplast DNA PCR were directed towards tmlF (cp5). (cp5F – ggttcaagtccctctatccc, cp5R
(GenBank Accession numbers AGK1943-AGK2163, JN400291-JN400323). Electronic
supplementary material: 3: Sequence. [2]).
The 220 DNA samples used in this study were collected and extracted by JDP and CEV in 2008.
They were also the basis of two published studies which assumed that the Boab mutation rate
was the same as the African rate, an assumption leading to a discrepancy that was explained by
proposing a “Pleistocene refugium” (Bell et al 2014, Rangan et al 2015 ). The present study was
stimulated by the delayed realisation that African and Australian mutation rates might not be the
same.

PCR and Sequencing: Primers ITS.LEU (gtcactgaaccttacatttag) and ITS4 (tcctcgettattgatagc)
were used to amplify and sequence the nuclear ribosomal internal transcribed spacer region
(nrITS) (GenBank Accession numbers AGK1943-AGK2163, Electronic supplementary material:
3: Sequence. Typical concentrations were 3-10 ng/µl. The PCR product was used as template for
sequencing in a 10 µl reaction volume using 1.75 µl Big Dye Buffer, 0.5 µl Big Dye v3.1, 1.5 µl
purified PCR product, and 0.32 µm primer. The labelled product was purified as described above
and sequenced at the Australian Genome Research Facility (AGRF; Brisbane, Australia).
Resulting sequences were aligned using ClustalW
(http://www.ch.embnet.org/software/ClustalW.html).

Phylogenetic Analysis. --- Multiple sequence alignments were carried out using MUSCLE,
through the Geneious interface, with eight iterations. Regions where the alignment was
ambiguous were removed, indels of more than one nucleotide were re-coded as a single
character, and then analyses were run with gaps treated as a new character state, and with gaps
treated as unknown. Parsimony analyses were carried out in PAUP* 4.0 β 10 (Swofford, 2002).
For each DNA sequence region, a heuristic search with 1000 replicates, random addition of taxa, and tree bisection reconnection (TBR) branch swapping was used. Bootstrap analyses were carried out with 1000 bootstrap replicates, each with 100 heuristic search replicates, with random addition of taxa and TBR branch swapping. Uncertainty about the topology of the basal Australian, African and Malagasy lineages was reduced by treating gaps as a new character state (as described above), reducing the number of sequences in the large Malagasy clade, and checking different outgroups (*Cavanillesia* was most effective in reducing the number of trees), with the result that the number of equally most parsimonious trees was reduced from 30 to 12 for analysis using ITS sequence, all of which favour a linkage between the Australian and African clades, with the Madagascan clade at the basal node.

Bayesian analysis was carried out using the MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) plugin version 2.0.3 for Genious (Biomatters Ltd.). The optimal models of sequence evolution for each gene region were determined with the Akaike Information Criterion (AIC) using Modeltest, through the Geneious interface. Using the selected model, two simultaneous searches were performed, each with four MCMC chains (one cold and three hot), with a sample frequency of 1000. Every 1,000,000 generations, the standard deviation of split frequencies between the two simultaneous analyses was calculated and when this decreased to less than 0.01 the search was stopped. This process was repeated with two further simultaneous runs of the MCMC, and the results were compared to confirm that apparent convergence of the first two runs wasn’t due to both reaching the same local optimum. Plots of the likelihoods of sampled trees were examined to determine when the MCMC chains had reached stationarity, and the sampled trees prior to this were discarded as burn-in. A majority rule consensus tree was obtained from the remaining trees.

Rationale for Comparison of Parsimony and Bayesian Analyses:

We compared the results of phylogenetic reconstruction using PAUP and Mr. BAYES using the same starting matrix, one that had been shown previously to give a relatively stable topology in which the tetraploid is in the crown and the Boab is closer to the African baobabs than it is to the Malagasy baobabs. The parsimony method provided direct access to a phylogeny in which branch lengths were proportional to mutation rate, without the complications of rate adjustment.
that are implicit in Bayesian reconstructions and without the need to provide a specific quantitative hypothesis about the change in rate that underlies the a posteriori Bayesian approach. The role of the Bayesian method was to provide a phylogenetically appropriate context for the relative rates of individual taxa in PAUP that could be used to estimate the absolute magnitude of each. There were no clues from the natural history or phylogeny that might help one to guess if the Boab had an altered mutation rate or what the altered rate might be. This made it difficult to frame a hypothesis for a posterior Bayesian testing. On the other hand, the flexibility of this method makes it appropriate to provide the phylogenetic context for each rate. Of particular importance here would be the estimated relative distance of A. gregorii from the African and Madagascan clades, a divergence which strongly affects the calculated rate of the A. gregorii clade (as shown in Fig. 1) and which is more finely estimated by Mr. Bayes than by PAUP.

Estimation of Boab nrITS Mutation Rate:

At the outset of the study, we presumed that the published, calibrated mutation rates for nrITS of the African Baobab, A. digitata, would also be applicable to its close Australian relative, A. gregorii. Using the African mutation rate gave very large estimates of Boab age, much larger than expected, suggesting that a more appropriate age would be obtained if we could use a more appropriate, larger, Australian mutation rate. The dangers inherent in our early assumption that A. gregorii would have the same mutation rate as A. digitata are apparent in the warning published by Kay et al. (2006) in their description of the 50-fold variation seen in independently calibrated nrITS rates from 28 plant taxa. "No significant phylogenetic signal to the rates was detected (p =0.393), suggesting that phylogenetic relatedness is not an appropriate justification when choosing rates from the literature." (Kay et al. 2006). Following this admonition, it would be reasonable to question whether the mutation rate of any of the other 8 baobabs is similar to the African rates. In fact, only A. gregorii has an obviously elevated rate in the relative rate tests, the time-dependent calculations, and the phylogenetic reconstructions, in keeping with its mysterious transport, its disjunct status, and its completely different, vacant, new niche in Australia’s Kimberley, compared to the other baobabs.

Supporting a divergent Australian mutation rate, we found large discrepancies between the age estimates given by nrITS substitutions using the African mutation rate, and other estimates such as those derived from microsatellites, or from the time to reach the current distribution from
cumulative dispersal of Boab seed (Electronic supplementary material: 1 : Macropod dispersal and Microsatellites). Although these other estimates were less precise, they were yet more than an order of magnitude lower than the estimate provided by the nrITS substitutions combined with the African mutation rate (Table 1).

**Fig. 3:** Eighteen-fold increase in branch lengths for all eight samples of nITS from A. gregorii in parsimony reconstruction by PAUP (on right). Rate calculations are dependent upon the phylogenetic topology, which is explicit in this PAUP reconstruction that is known to be congruent with other detailed work supporting the sister relationship between A. gregorii and the African baobabs. The parsimony reconstruction has transparency of the effects of rate that require more complex intervention in the less transparent a posteriori methodology of Bayesian analysis which provides a contextual baseline of rates across the genus (left side). Except for A. digitata, no revision of mutation rate has been attempted for any other of the remaining 8 baobabs, all of which could conceivably have a different mutation rate. These data show that A. gregorii has a rate that is different enough to be obvious, at 18X, compared to all other baobabs.
As a result, we thought that we may have underestimated the rate, a conjecture that was supported quantitatively by comparing reconstructions of the phylogeny of all 9 baobab species using "fixed clock" (PAUP conditions versus "relaxed clock" (MrBayes 3.2, [10]). This comparison is shown in Fig. 1, adapted from Pettigrew et al [2]; [Fig. S7C, PAUP vs Fig. S7F, MrBayes]. The Boab, *A. gregorii*, was the only baobab species of the 9 in the genus whose rate departed significantly in every sample from all other baobab taxa. Branch lengths of the Boab were on average 18.3x longer. There were 8 separate samples of DNA from different trees of *A. gregorii*, that each gave 720 bp of nrITS sequence. Lengths for these 8 *A. gregorii* branches on the Bayesian reconstruction of the phylogeny of all *Adansonia* species (Fig. S7F, [2]) were 0.07, 0.08, 0.068, 0.066, 0.1, 0.067, 0.069 substitutions, while the same 8 samples in the PAUP reconstruction gave branch lengths of 70.6, 8.2, 18.9, 12.7, 10.9, 3.8, 11.9 and 9.8 substitutions respectively (Fig.1), giving a mean ratio between the two methods of 18.3x +/- 20.2 S.D. Since both the PAUP and Bayesian trees had assumed a mutation rate of 2.48 x 10^-9 substitutions/site/year, the appropriate rate may therefore be 18.3X greater, at around 4.5x10^-9 substitutions/site/year. This is the “Australian” rate that has been used in this paper to calculate Boab age from nrITS substitutions (Table 1). This is a large mutation rate, but must be seen in the context of the very small number of taxa with very short calibration times, all of which also have very high mutation rates. The total range (50X) of published nITS rates (Kay et al. 2006).

Using the published time-dependence curves, one can estimate the change in the mutation rates of well-known taxa if their calibration time were extended slightly. (e.g. *Gossypium* would have the even greater mutation rate of 5.5.x 10^-8 substitutions/site/year if it were measured at 70 Ka instead of the established 8.5 Ma).

Mathematical Test of the Mutation Rate Increase Using Time-Dependence::

Explicit determinations of time dependence of mutation rate, like the curves delineated for time and mt DNA of humans, are not available for plants. Nevertheless, a cross check is possible for plant nITS from the survey of Kay et al (2006), which shows a close fit. For example, the plant taxon with the most recent calibration time of the 29 that were sampled (0.1 Mya), *Gentiana var Ciminalis*, also has the highest mutation rate. At 19 x 10^-9 this is 50x the rate of the lower taxa and 7x the rate of the mean. This species is therefore located on the steepest part of the time dependence curve that would be expected to give a rate around 20 if the calibration time were
shortened even further, from 0.1 Mya to 0.07 Mya (as indicated by nITS substitutions and the mutation rate calculated from the phylogeny of *Adansonia*) (MRPA).

To test our estimate of the magnitude of the increased mutation rate, we used the mathematical, time-dependence curves for mutation rate (Ho et al. 2005). These curves give relative values, but not the specific values for calibration time, or rate, of individual taxa, so we used *Gentiana var. Ciminalis* as a model because it has the shortest calibration time (100 Ka) and the highest mutation rate (19 x 10^{-9} substitutions/site/year) for any published nITS gene, in line with our heterodox proposal for an increased mutation rate in *A. gregorii*. We asked “from the curves”, what would be the consequences for this model taxon, if the calibration time were shortened even further, for example, to ~70 Ka instead of 1 Ma? The curves show that this would lead to an increase of mutation rate by a factor of 2.4X, to 4.6 x 10^{-8}, which is very close to the estimate of 4.5 x 10^{-8} for *A. gregorii* that we derived independently from phylogenetic analyses of the sequence data themselves. The close similarity between the rate derived from time-dependence and the rate derived from phylogenetic analysis suggests that time dependence alone may account completely for the raised mutation rate of a very recent species, such as *A. gregorii*. Note that this test is guided by the hypothesis that very short calibration times and high mutation rates might be involved, but that these surprisingly consistent results from the mathematics are independent of those arising from the comparison between different phylogenetic reconstructions and patterns of substitution (Fig. 1). The difference between the two analyses is that in the first case we have used phylogeny of the nITS substitutions with different clock constraints to estimate a mutation rate which could then be used to arrive at a date. In the second case, we began with hypothetical dates and then derived mutation rate using the time dependence curves.

In addition to analysis based on all the substitutions in the 220x720 nITS matrix, an alternative analysis was based on biogeographical information which enabled a comparison of "Old" trees from the NW with "New" (EK) trees from the Eastern Victoria R. that had been defined by microsatellites ([6], Electronic supplementary material 1: Macropod dispersal and microsatellites]. Although nITS substitutions had not permitted the prior identification of the sub-regions identified by microsatellites, the different pattern of nITS substitutions in the East and West sub-regions enabled a separate calculation of age; 71 Ka (Table 1)
Results and Discussion:

Relative Rate Test:
The Classical method for assessing mutation rate variations is the relative rate test, which is shown in Fig. 1 for the Boab. This test is incorporated into some phylogenetic reconstructions such as Mr. Bayes, but this figure shows why we did not use this approach. The problem of dependence of rate upon the topology of the phylogeny means that derivation of rate may not be so obvious when hidden inside an algorithm like Mr. Bayes. The exercise showed that the Boab has an elevated rate compared with other baobabs and that the elevation is greatest when the node giving rise to the Boab is close to the African baobabs. When the node of Boab branch is close to the Madagascan baobabs the rate increase is 2X, but when the node is close to the African clade, the calculated increase in rate may be as much as 18X.

Time-Dependence of Mutation Rate:
The curves for time dependence are all very similar, but there is no scale for mutation rate on the ordinate, so we tested a variety of recent times to check them against mutation rate. This gave a remarkable correspondence between the mutation rate that we estimated from phylogenetic reconstruction and the mutation rate given by the time-dependence curve at 70 ka. The correspondence is remarkable because there is a large number of possible combinations of mutation rate, nITS substitutions and coalescence time, from which the curves seem to have chosen the appropriate one.

nITS Substitutions: There were 566 steps in the consensus of trees generated by PAUP from the matrix of 220 ITS sequences (Electronic supplementary material: 3 : Sequence; GenBank Accession numbers AGK1943-AGK2163, JN400291-JN400323 ). All phylogenetic trees generated from the population of 220 individuals of A. gregorii were monophyletic. The substitutions were evenly distributed throughout the matrix and so were consistent with clock-like behaviour. Using the estimated Australian mutation rate for A. gregorii of $4.5 \times 10^{-8}$, we arrive at timing for this baobab in Australia of ~70 Ka. Using the African mutation rate instead, timing is estimated at 1200 Ka (Table 1). Biogeographical information from microsatellites studied in these same DNA samples
allowed the detection of a contrasting set of substitutions in western vs, eastern populations. There was a difference of 20 substitutions between 7 trees in sub-group NC (“Old”) on the Northwest Coast and 2 EK trees (“New”) in the East Kimberley. The divergence of 0.38% gives a coalescence timing of 1234 Ka (African rate) or 88 Ka (Australian estimate) (Table 1). The information from chloroplast haplotypes was not useful in the present studies (see Pettigrew et al. 2012) except for a finding of the same, unusual, cp5 haplotype in two widely-separated trees on the NW and E sides of the Kimberley plateau. As well as being widely separated, by 180 km of Kimberley coastline, they were not connected to any trees with the same haplotype on the plateau itself, so must have been the result of dispersal on the floodplain, around the NW to NE coastal margins of the plateau. These “floodplain species” were part of the same population of N and NC trees that were already shown to be “young” by both DNA microsatellites (Bell et al. 2014) and nITS substitutions (ibid).

“Australian” mutation Rate: Phylogenetic analysis of all 9 baobab species with a fixed clock (PAUP) revealed that the Boab is an outlier with much longer branch lengths on every tree sampled. These increases amount to a 18x increase in the rate of evolution when compared to the other species, or to the same samples of Boabs in a similar reconstruction using a “relaxed clock” (Fig. 3). Since the base mutation rates used for these reconstructions were the published values of the rate for *Adansonia digitata* of 2.48 - 2.9 x 10^-9 substitutions/site/year, the estimated final mutation rate for the Boab is 4.5 – 5.2 x 10^-9 substitutions/site/year, which gives a timing in Australia of 69 – 78 Ka (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>ASSUMED</th>
<th>MEASURED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutation Rate</td>
<td>2.5 X 10^-9</td>
<td>45 x 10^-9</td>
</tr>
<tr>
<td>Basis for Rate</td>
<td>African Rate</td>
<td>Analysis (9 baobabs)</td>
</tr>
<tr>
<td>Divergence (<em>A. digitata-A. gregorii</em>)</td>
<td>0.6 %</td>
<td>0.6 %</td>
</tr>
<tr>
<td>Divergence Time</td>
<td>2.4 Mya</td>
<td>72 Kya</td>
</tr>
</tbody>
</table>
Coalescence Time vs. Divergence Time:: There are now many more African/Australian alignments available from nITS sequence compared with the single Australian sequence that provided the divergence time to the stem node by Baum et al (1998). The new sequences of the present study give a divergence to the African diploid, *A. kilima*, of 0.6%, or 2.4 Ma if one makes the assumption of equality of rate between African and Australian baobabs (2.5 X 10$^{-9}$ substitutions/site/year). By convention, the stem node is midway, at 1.2 Ma, but note that this also assumes a low mutation rate for the *A. gregorii* limb of the divergence that would have to be revised upward by the measures here. If one uses the revised, high rate (45 X 10$^{-8}$ substitutions/site/year), derived from two independent calculations, the time for the limb on the *A. gregorii* side of the stem node becomes 68 Ka. In other words, if one uses the new estimate of mutation rate for *A. gregorii*, coalescence time (67 Ka) and divergence time (68 Ka) are very close. This hypothetical date for the divergence time can be further adjusted, because of its unknowns, compared to coalescence time. For example, transport to Australia may not have occurred precisely at the time when the precursor split from the stem. The resulting “waiting time” is another unknown with respect to mutation rate and timing. There are many resulting scenarios. Note that these do not show any superiority of divergence time over coalescence time. Rather, the two measures generally support each other at~70 Ka, with small adjustments of the “waiting time” and heterogeneous mutation rates across the node allowing a perfect match between coalescence and divergence times. The close similarity between coalescence time and divergence time in calculations using the revised Australian rate does not leave any spare time for the postulated Pleistocene refugium and bottleneck in the Kimberley. The similarity of these two timing measures also underlines the validity of coalescence time, as well as divergence time, as a measure of the Boab’s arrival in Australia.

Two previous studies raised the possibility that the Boab might have been a more recent arrival than generally believed.

1. Limited NE Distribution: Bowman (1997) was very familiar with the Boab distribution and was
puzzled that the active dispersal front on the NE flank of the Boab distribution had not yet taken Boabs into the Gulf region. There was no obvious restriction of the NE expansion of the distribution, such as fire, that had been studied and eliminated. An obvious explanation for the puzzling observation of this limit to the distribution is that the Boabs were much more recent than expected.

2. Young vs Old Controversy: A recent study of Boab genetics using microsatellites showed that Boabs appeared to have a recent origin. The young microsatellite dates and the low genetic diversity were in conflict with the older 2.4 Ma estimate (1.4 Ma in this study: Table 1) using the African mutation rate. Rather than testing a possible alternate rate, the solution adopted to deal with this discrepancy was the invention of a "Pleistocene refugium" from which modern, young Boabs had sprung without revealing any evidence of their older, precursor stock. There is no experimental support for this invention, by definition, which owes its existence solely to the conflict between the "young genetics" of the microsatellite data and the "old inference" from the African mutation rate. If the Boab is actually recent, as indicated by a higher Australian rate, and the assumption of the African rate is incorrect, there is no discrepancy and the need for the "Pleistocene refugium" disappears. The recently-extended proposal, of a Pleistocene refugium on the Timor floodplain before the inundation of the Holocene (Rangan et al. 2015), also has the problem that there are no obvious genetic remnants of it. There is good evidence that there were Boabs on the floodplain in the form of cp5 haplotypes on either side of the Kimberley peninsula that were not connected to any trees with the same haplotype on the plateau and must therefore have been dispersed on the floodplain before inundation (ibid). While this is evidence for the presence of Boabs on the floodplain before inundation, it is not evidence for a refugium, as none of these floodplain trees has the predicted Pleistocene age, belonging instead to the recent N and NW microsatellite subgroups (Bell et al. 2014). A floodplain refugium may also be a dubious concept because of the numerous breaks in the Kimberley escarpment that would have allowed floodplain Boabs to reach the plateau via one of their favourite riverine dispersal routes, provided by the outlets of the Fitzroy, Ord, Victoria and Fitzmaurice Rivers, whose present day Boabs all have relatively the young ages found in the microsatellite DNA study (Bell et al. 2014).

? Sufficient Time for a New Species.
The estimated Australian mutation rate gives a short coalescence time for A. gregorii in Australia
that raises a question in the minds of some observers about whether 70 Ka is sufficient time for the evolution of a new species. A number of points address this issue:

1. At a rate 18X that of some other trees, the species-specific changes will accumulate faster than judged by absolute time scales.

2. *A. gregorii* is a "good" species of baobab with distinctive floral evolution that includes an upright bundle of anthers, compared with the pendant anthers of its African sister taxa, and the more loosely oriented anthers of the Madagascan baobabs (Pettigrew et al 2012, Baum et al 1998). This major change in the orientation of the Boab anthers could have been accomplished despite the small absolute time interval, given the intense selective pressure of the main pollinator, *Pteropus alecto*, which consumes the anthers completely from above while it is perched beside the flower ("G. Rethus and J. Pettigrew" in prep). This large flying fox is 10x the size of the small megabats (*Roussetus, Epomophorus*) which pollinate the pendant African baobab flowers while hanging from them, an option not available to *P. alecto* that may help explain the evolution of the solid attachment and upright anthers of the Boab flower.

3. There are numerous examples of rapid plant speciation such as: a. hexaploid wheat and teosinte maize that have appeared as part of the beginnings of human agriculture around 10 Ka ago (Ranere and Piperno 2009); b. new neotropical forest species that have appeared in 50 Ka (Rull 2008), as well as new species of daisies (Ungerer et al. 1998); c. the numerous new species of *Corus* that have evolved recently as part of their pollination syndromes with hummingbirds (Kay et al 2005).

*Extreme nITS Mutation Rate:*

While all five measures indicate an increase in mutation rate, the value given by phylogenetic analysis of nITS substitutions is 4.5 x 10-8 substitutions/site/year. This is something of a record, with the nearest published rate for this gene being in *Gentiana var. Ciminalis* at 19x10-9 substitutions/site/year, a plant taxon which is also distinguished for its very short calibration time (0.1 Ma). To put this into context, there is a 50x variation in the published mutation rate for this gene (Kay et al. 2006). Moreover, the published distribution of nITS rates is biased toward longer calibration times, so one might expect more extreme rates like those of Boab and *Gentiana var. Ciminalis* if more taxa with short calibration times are studied. *A. gregorii* has a very short calibration time with which the high mutation rate is quantitatively consistent, based on time-dependence curves. *A. gregorii* may have a record rate at present, but this is well contained within
the range of plant nuclear rates (3-50 x 10-9/site/year) (Wolfe et al. 1987).

Conclusion: Five different lines of evidence support an inequality between Australian and African baobab mutation rates. Each indicates that the change in the mutation rate of A. gregorii is an increase. Two measures are quantitative and agree that the increase in mutation rate is 18X.

1. Time Dependence: Even if one uses the African rate that is under question, the calculated divergence time is nominal at 1.4-2.4 Ma depending on the study, values inside the calibration times that are known to be associated with increased rates because of time dependence. These nominal values of 1.4-2.4 Ma cannot therefore be taken at face value because we do not know the relative contributions to it of African and Australian limbs of the phylogeny, with the likely contribution of the Australian limb having many more substitutions as a result of the increased rate.

2. The Classical Relative Rate Test of Yang shows an obvious increase in the A. gregorii rate whose magnitude increases markedly in phylogenetic reconstructions where A. gregorii branches closer to the African baobabs than to the Malagasy baobabs, a topology that is favoured in recent studies. In addition to showing that A. gregorii has an increased rate, this test emphasises the importance of the particular phylogenetic reconstruction used in the relative rate test.

3. Comparison of parsimony and Bayesian reconstructions allowed the increase in mutation rate to be quantified accurately. The indicated rate is very high, but is still in keeping with the 50x range of published rates for this gene. While it is quite close to the published record rate for this gene, it is also associated with the very short calibration times that are elsewhere associated with very high rates.

4. Time-dependence curves provide a highly quantitative relationship between rate and calibration time. These curves predicted the same rate that we obtained in the phylogenetic reconstructions using information about substitutions in nITS.

5. Finally, there is a number of observations where there is a discrepancy between the apparent young age of the Boab and the much older age predicted from lower, African rate. These include microsatellite DNA studies of the Boab (Bell et al. 2014), and estimated dispersal across the Kimberley by the main macropod agents. Both of these estimates are very recent, less that 100 ka, neither of which are easy to reconcile with the 1.4 - 2.4 Ma given by the African rate, but perfectly in keeping with the younger age given by the increased mutation rate.
Conclusion:
This study provides no support for the commonly-made assumption that the mutation rate for the nITS gene of the Boab is the same as its African sister taxon, *A. digitata*. In the largest survey of nITS mutation rate so far, Kay et al (2006) also cautioned against this assumption, because close relatives can have wildly different rates. Instead, three approaches, the classic relative rate test, time dependence of mutation rate and phylogenetic analysis, all indicate a greatly increased mutation rate for the Boab, with the most quantitative method giving a 18x increase. This is large, but still within the 50x range of mutation rates of nITS. Combined with the nITS substitutions revealed in this study, an 18x increase in rate gives an arrival time in Australia for the Boab around 70 Ka, which is also close to the time of the Toba mega-eruption (Storey et al. 2012) that had a big impact on modern humans (Robock et al 2009), many of whom subsequently left Africa, the source of the Boab's progenitor (Rasmussen et al. 2011). The extreme utility of baobabs (Supplementary Material) would recommend them to migrants, so human transport might explain the very narrow distribution of Boabs in the Kimberley, much narrower than expected if seed arrived by flotation, as well as the completely overlapping distribution of rock art that has numerous depictions of Boabs (Pettigrew 2012). A link between the age of Boabs in Australia from molecular genetics, and the age of depicted Boabs in rock art, is presently only circumstantial, but might become substantive with advances in the dating of individual samples of rock art.

The best estimate from this study for the timing of the Boab's arrival in Australia is around 70 Ka (Table 1). This happens to coincide with the Toba mega-eruption (Robock et al 2009; Storey et al. 2012) which is thought to have triggered the first modern human migrants to leave Africa (Endicott et al. 2009).

Acknowledgements:
Claudia Vickers played an essential role in refinement of the technique for DNA extraction from Boab leaves and the subsequent preparation for DNA sequencing. Chloe Callistemon assisted with leaf collection. Lars Nielsen made available his excellent laboratory facilities in the *Australian Institute for Bioengineering and Nanotechnology*. Lance Conley at *Lone Eagle Aviation* (Kununurra WA) provided essential support for helicopter-based collections in difficult locations.

Deleted: Australian Genome Research Facility (AGRF) at the University of Queensland generously donated all the sequencing required to complete this project.
All the Sanger sequencing of the nITS gene in this project was generously donated by the
Australian Genome Research Facility (AGRF) at the University of Queensland;

References:


Baum DA, Oginuma K. 1994. A review of chromosome numbers in Bombacaceae with new
counts for *Adansonia*. *Taxon* 43, 11–20

baobabs (Adansonia, Bombacaceae) as inferred from multiple data sets. *Systematic Biology* 47,
181–207. http://dx.doi.org/10.1080/106351598260879

Genetic diversity and biogeography of the Boab *Adansonia gregorii* (Malvaceae: Bombacoideae)
*Australian Journal of Botany* 62(2), 164-174 http://dx.doi.org/10.1071/BT13209

Bowman DMJS (1997) Observations on the demography of the Australian
Boab (*Adansonia gregorii*) in the north-west of the Northern Territory,


Reich, David; Green, Richard E.; Kircher, Martin; Krause, Johannes; Patterson, Nick; Durand, Eric Y.; Viola, Bence; Briggs, Adrian W. & Stenzel, Udo (2010), Genetic history of an archaic hominin group from Denisova Cave in Siberia, *Nature* 468 (7327), 1053–1060, doi:10.1038/nature09710, PMID 21179161.


http://dx.doi.org/10.2307/2694252


