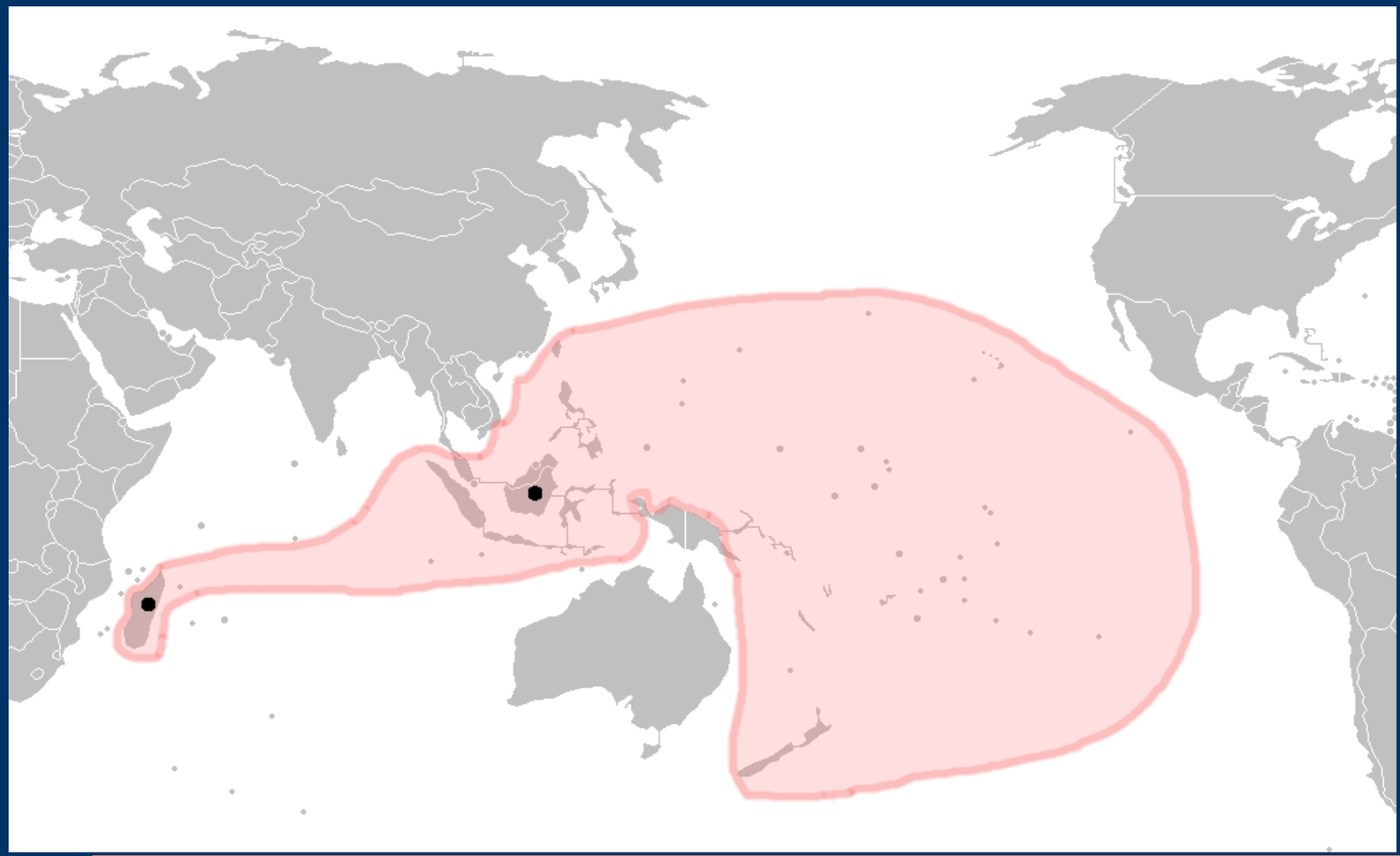
The phylogeny of Malagasy dialects

- Maurizio Serva
- Philippe Blanchard
- Eric W. Holman
- Filippo Petroni
- Dimitri Volchenkov
- Søren Wichmann

- DNA of Malagasy people is 50% African and 50% Indonesian, nevertheless all Malagasy dialects belong to the Western Indonesian subgroup of Austronesian family.
- The Indonesian language most closely related to Malagasy is probably Maanyan of South-east Kalimantan with a 45% of shared basic vocabulary, but close languages can be found in Sulawesi, Malaysia and Philippine Islands.



- Austronesian geography

Goal

The history of Madagascar peopling and settlement is subject to dispute and alternative interpretations among scholars. It seems that Indonesian sailors reached Madagascar by a maritime trek at a time which is disputed and can be between one and two thousand years ago, furthermore it is not clear if there were multiple settlements during centuries or a single one. More mystery is added by the fact that Maanyan, which seems to be the closest language to Malagasy, is spoken by a population which lives along the rivers of Kalimantan and which does not possess the necessary skill for long maritime navigation. A possible explanation is that they arrived with Malay sailors as slaves and both peoples are the ancestors of modern Malagasy. In this case, the dialects should show both a Malay and a Maanyan contribution.

In our research we would like to try to address these issues by a glottochronological analysis based on the automated method recently proposed by the authors (Serva and Petroni, Holman *et al.* , Petroni and Serva, Bakker *et al.* , Blanchard *et al.*).

Data and method

The data. collected by one of us (M.S.) at the beginning of 2010 with the invaluable help of Joselinà Soafara Néré, consist of a vocabulary of 200 words for 23 dialects covering all areas of the Island, probably the largest collection of comparative Swadesh lists for Malagasy variants.

The automated method uses a normalized Levenshtein distance to compute distance between words with same meaning and it averages over all the words contained in the list to obtain the lexical distances between pairs of dialects or languages. Then, these lexical distances are transformed, by a simple rule, into separation times (genealogical distances) .

There are some differences between the method by Petroni and Serva and the method by Holman *et al*. , Bakker *et al*. concerning the length of the lists and the normalization but in this talk we will merge the two points of view.

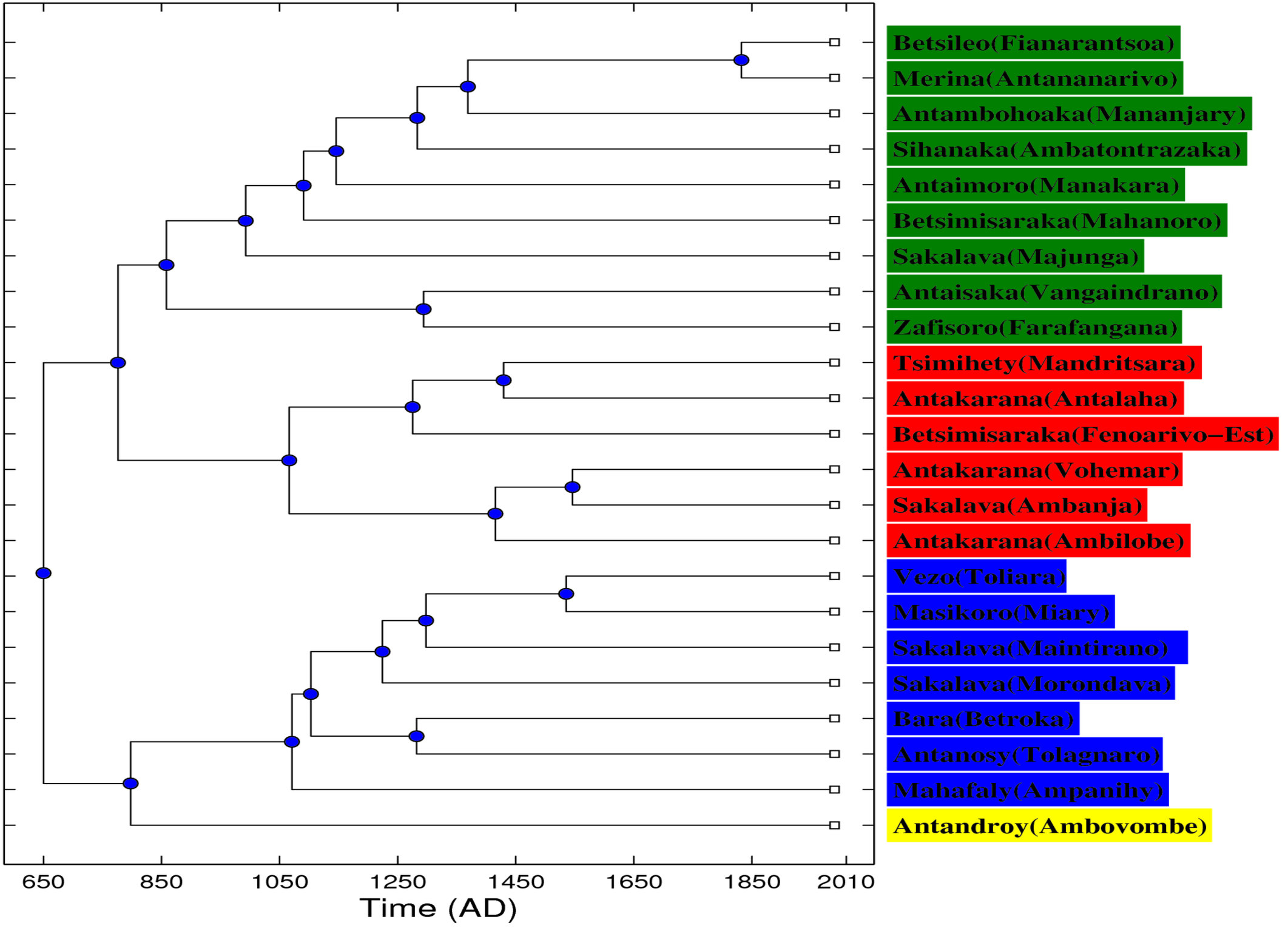

AMBILOBE
ANTALAHA
MAJUNGA
MANDRITSARA
FENOARIVO
MAINTIRANO
ANTANANARIVO
MAHANORO
MORONDAVA
MANANJARY
MANAKARA
FARAFANGANA
BETROKA
AMPANIHY

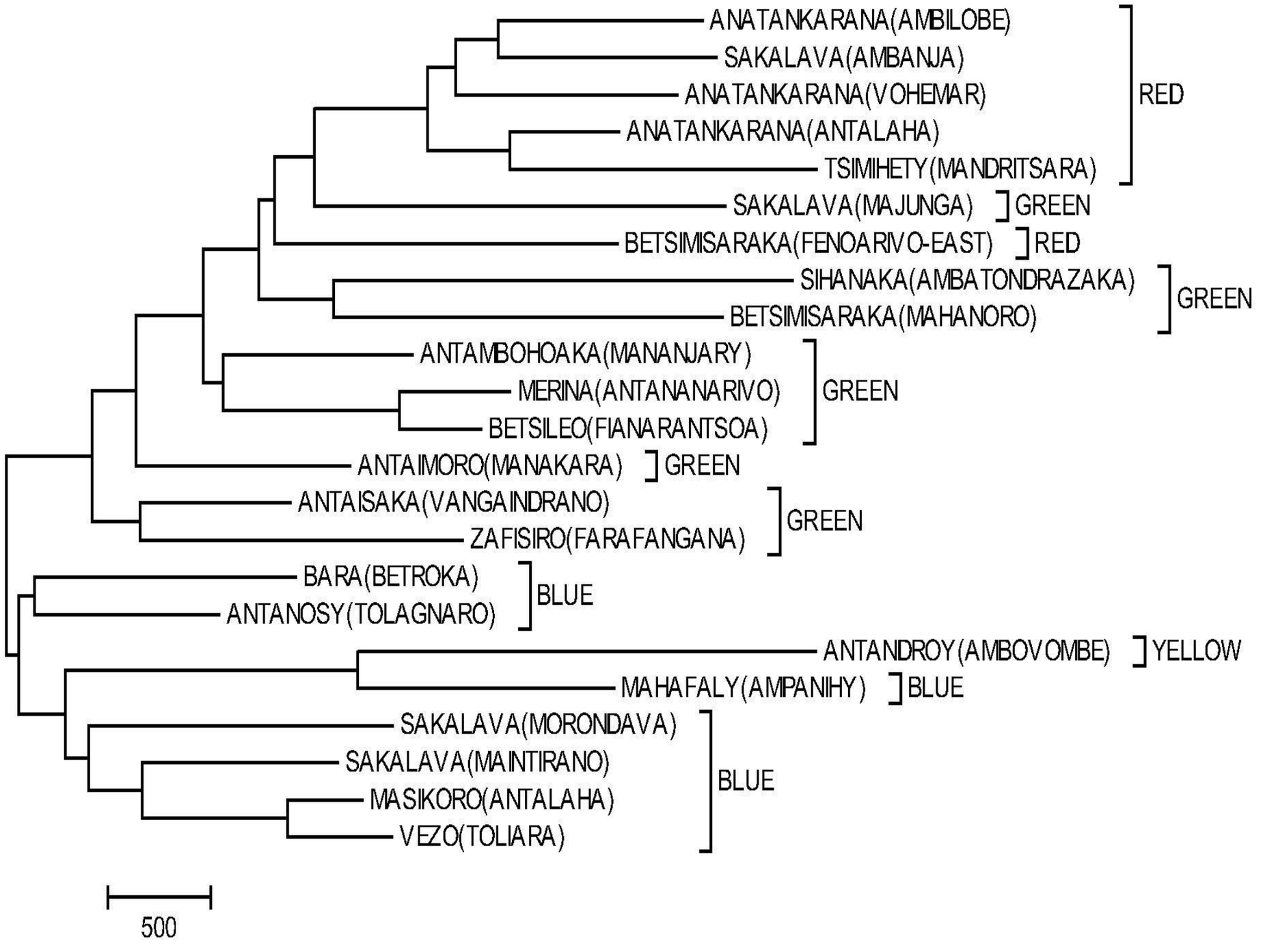
We computed the distance between pairs of Malagasy dialects and between Malagasy dialects and other Austronesian languages (like Maanyan). In the first case we obtain results concerning the internal relationships which can give information about dates and modalities of the settlement. In the second case we obtain results concerning external relationships, which can give information concerning the Indonesian origin of Malagasy ancestors.

In the next three slides we show how these distances can be handled in order to extract useful information from their tree-like organization.

In the next slide we show the result obtained using the Petroni and Serva approach and UPGMA method for three reconstruction. Then, we use the Holman *et al*. , Bakker *et al*. approach and Neighbor-joining method for the tree reconstruction. Finally, we show how results can be interpreted in terms of geography.

In both approaches results show that the 23 Malagasy dialects are divided into two main groups: south-west and east-center-north. This differs from the claims of Vérin *et al*. (1969) which propose a main division north/others.


650
850
1050
1250
1450
1650
1850
2010
Time (AD)
Betsileo(Fianarantsoa)
Merina(Antananarivo)
Antambohoaka(Mananjary)
Sihanaka(Ambatontrazaka)
Antaimoro(Manakara)
Betsimisaraka(Mahanoro)
Sakalava(Majunga)
Antaisaka(Vangaindrano)
Zafisoro(Farafangana)
Tsimihety(Mandritsara)
Antakarana(Antalaha)
Betsimisaraka(Fenoarivo−Est)
Antakarana(Vohemar)
Sakalava(Ambanja)
Antakarana(Ambilobe)
Vezo(Toliara)
Masikoro(Miary)
Sakalava(Maintirano)
Sakalava(Morondava)
Bara(Betroka)
Antanosy(Tolagnaro)
Mahafaly(Ampanihy)
Antandroy(Ambovombe)


ANATANKARANA(AMBILOBE)
SAKALAVA(AMBANJA)
ANATANKARANA(VOHEMAR)
ANATANKARANA(ANTALAHA)
TSIMIHETY(MANDRITSARA)
RED
SAKALAVA(MAJUNGA)
GREEN
BETSIMISARAKA(FENOARIVO-EAST)
RED
SIHANAKA(AMBATONDRAZAKA)
BETSIMISARAKA(MAHANORO)
GREEN
ANTAMBOHOAKA(MANANJARY)
MERINA(ANTANANARIVO)
BETSILEO(FIANARANTSOA)
GREEN
ANTAIMORO(MANAKARA)
GREEN
ANTAISAKA(VANGAINDRANO)
ZAFISIRO(FARAFANGANA)
GREEN
BARA(BETROKA)
ANTANOSY(TOLAGNARO)
BLUE
ANTANDROY(AMBOVOMBE)
YELLOW
MAHAFALY(AMPANIHY)
BLUE
SAKALAVA(MORONDAVA)
SAKALAVA(MAINTIRANO)
MASIKORO(ANTALAHA)
VEZO(TOLIARA)
BLUE
500

- In both trees the 23 Malagasy dialects are divided into two main groups: south-west and east-center-north.
- In the first tree, each of the two main groups is divided in two more subgroups corresponding to the geographical regions in the map on the right.
- The tree constructed using NJ and shorter lists of 40 words has exactly the same main partition but the south-west group has a different internal partition since the Antandroy dialect (Ambovombe) is grouped closer to the other southern dialects.
- In next slide we measure the average distance of each dialect to all the others, the largest distance is for Antandroy, the smallest for Merina which is the official language.

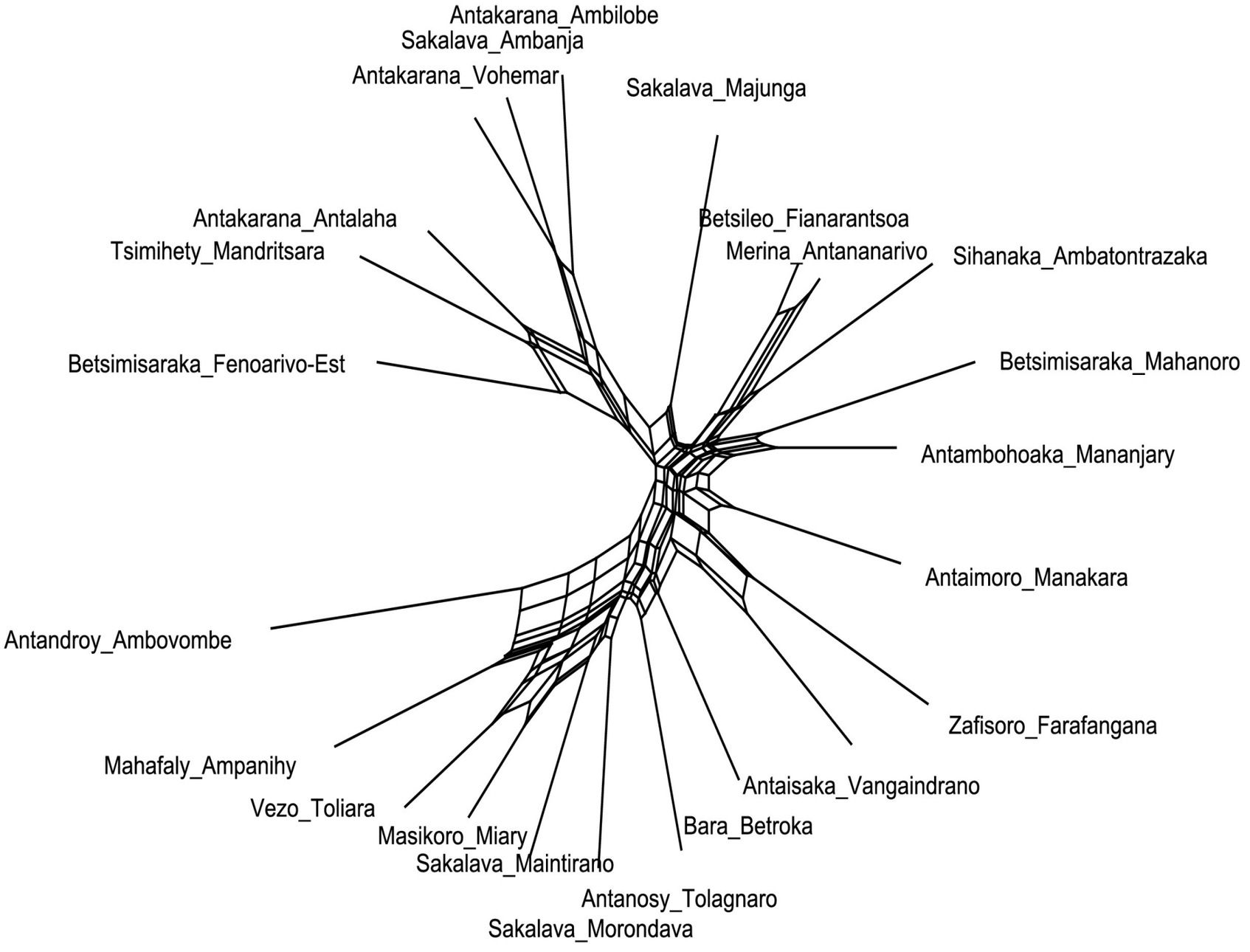
0.5
0.4
0.3
0.2
0.1
0
Antandroy(Ambovombe)
Mahafaly(Ampanihy)
Tsimihety(Mandritsara)
Zafisoro(Farafangana)
Betsimisaraka(Mahanoro)
Antakarana(Ambilobe)
Sihanaka(Ambatomitrazaka)
Vezo(Toliara)
Sakalava(Ambanja)
Antakarana(Vohemar)
Betsimisaraka(Maljunga)
Sakalava(Fenoarivo–Est)
Sakalava(Morondava)
Sakalava(Maintirano)
Antaisaka(Vangaindrano)
Masikoro(Miary)
Antanosy(Tolagnaro)
Antaimoro(Manakara)
Bara(Betroka)
Antambohoaka(Mananjary)
Betsileo(Fianarantsoa)
Merina(Antananarivo)

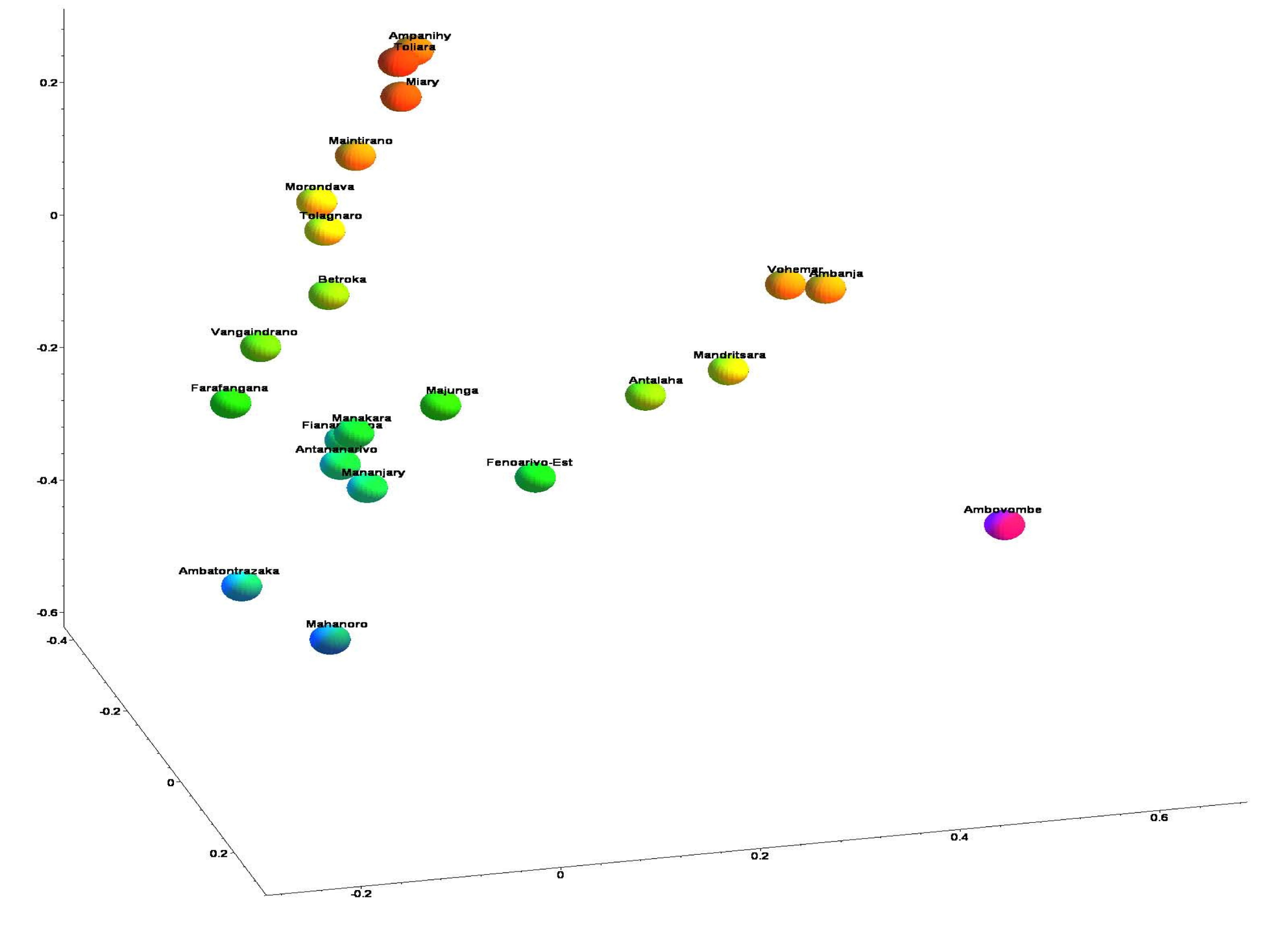
- The above result confirms that Antandroy is a quite deviant dialect as it is clearly perceived by other Malagasy peoples. Nevertheless, this difference is less relevant if one looks at core vocabolary.

- Probably a tree is not an appropriate description and it is more useful to perform a network analisys.

- In the next slide we show the output of NeighbourNet, using Splitstree 4. The geography of Madagascar is confirmed and Antandroy is scattered at one extreme.

- Structural Component Analysis (Blanchard *et al*.) is a powerful tool to find correlations of any order among related components. We have produced a three dimensional representation of the dialects that is shown in the slide after the next one.


Antakarana_Ambilobe
Sakalava_Ambanja
Antakarana_Vohemar
Sakalava_Majunga
Antakarana_Antalaha
Tsimihety_Mandritsara
Betsileo_Fianarantsoa
Merina_Antananarivo
Sihanaka_Ambatontrazaka
Betsimisaraka_Fenoarivo-Est
Betsimisaraka_Mahanoro
Antambohoaka_Mananjary
Antaimoro_Manakara
Antandroy_Ambovombe
Zafisoro_Farafangana
Mahafaly_Ampanihy
Antaisaka_Vangaindrano
Vezo_Toliara
Bara_Betroka
Masikoro_Miary
Sakalava_Maintirano
Antanosy_Tolagnaro
Sakalava_Morondava


Ampanihy
Toliara
Miary
Maintirano
Morondava
Tolagnaro
Betroka
Vangaindrano
Farafangana
Majunga
Manakara
Antananarivo
Mananjary
Fenoarivo-Est
Vohemar
Ambanja
Mandritsara
Antalaha
Ambovombe
Ambatontrazaka
Mahanoro
0.2
0
-0.2
-0.4
-0.6
-0.4
-0.2
0
0.2
-0.2
0
0.2
0.4
0.6

- The two linear components in the previous slide correspond to the North/South geography of Madagascar. Antandroy is scattered very far away.

- The distribution of dialects along the radial direction is remarkably heterogeneous indicating that the rate of changes in the orthographic realizations of Swadesh's vocabulary was anything but constant (which also implies that one of the UPGMA main assumptions is violated).

- The variance of the radial component has been computed and it is shown in next slide. Since this variance can be linearly associated to the time distance of contemporary dialects from the protolanguage, our results suggest that landing in Madagascar was around 600 A.D.

- The landing place was probably close to Makara and Manajary (but Majunga is also possible) since these two languages are at the origin of the two linear components. This result, as we will see later, is strongly confirmed by an analysis performed according to the method by Wichmann *et al.*.

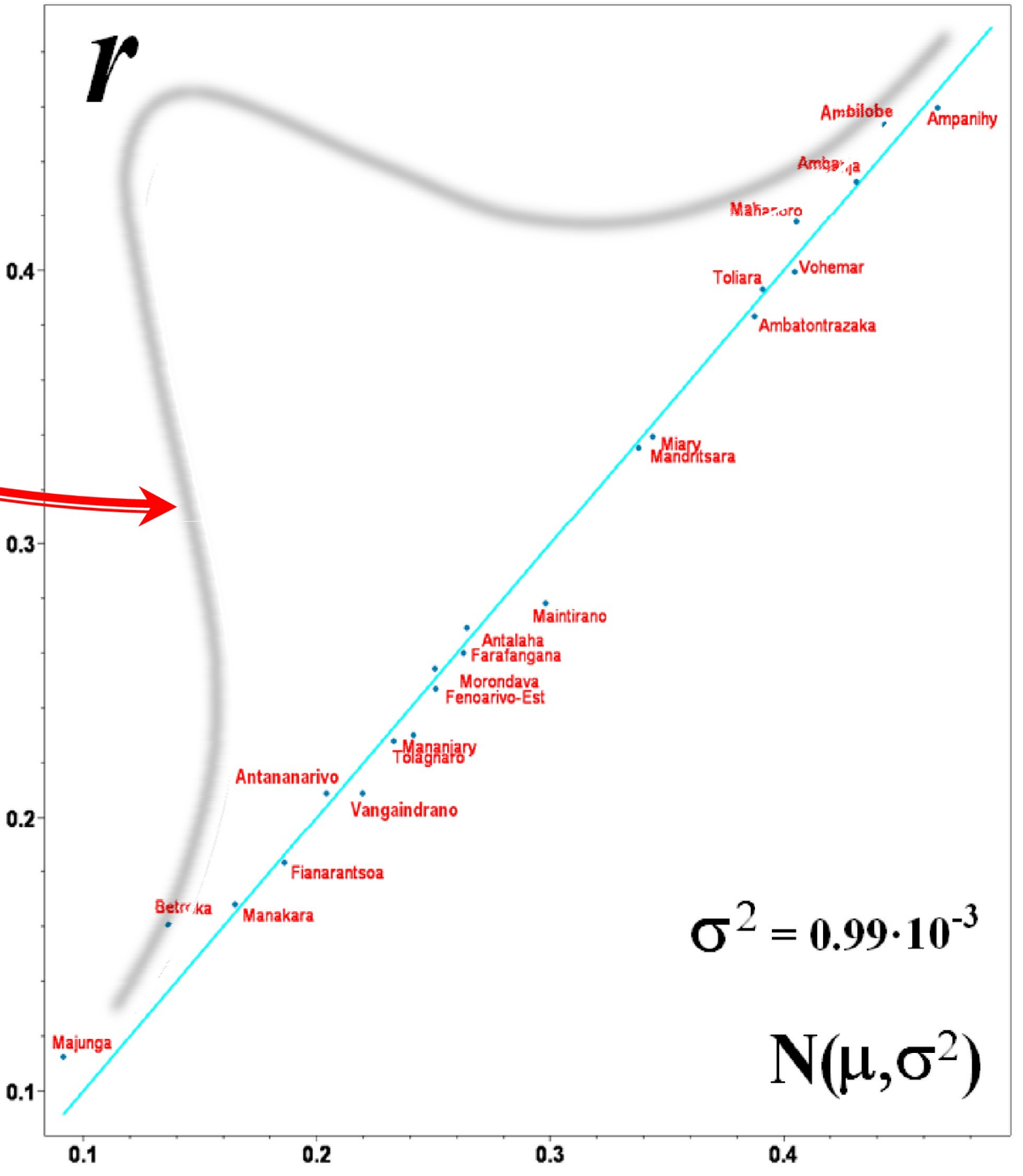
$$\frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \rightarrow \frac{\exp\left(-\frac{(x-\mu)^2}{2t}\right)}{\sqrt{2\pi t}}$$

$$\frac{t}{\sigma^2} = (1.367 \pm 0.002) \times 10^6$$

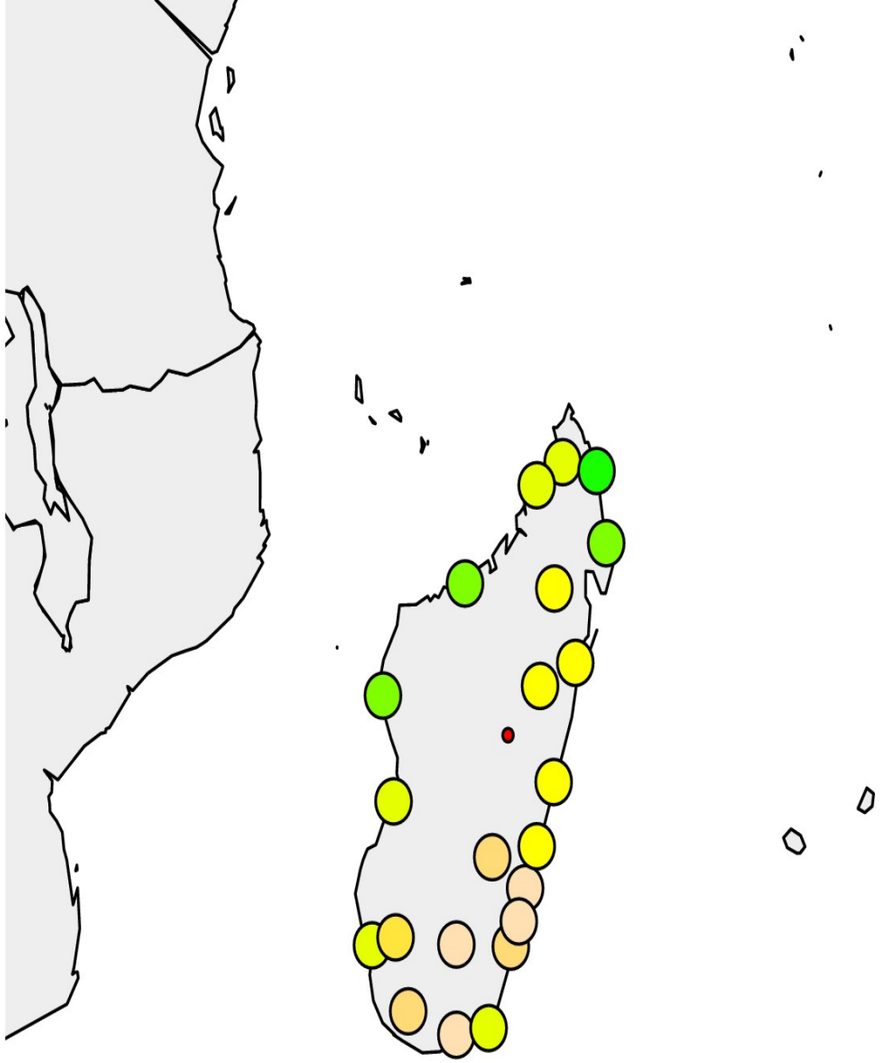
Blanchard, Petroni, Serva, Volchenkov (2010); doi:10.1016/j.csl.2010.05.003

t=1367

~ 2010 – 1367 = 643 A.D.

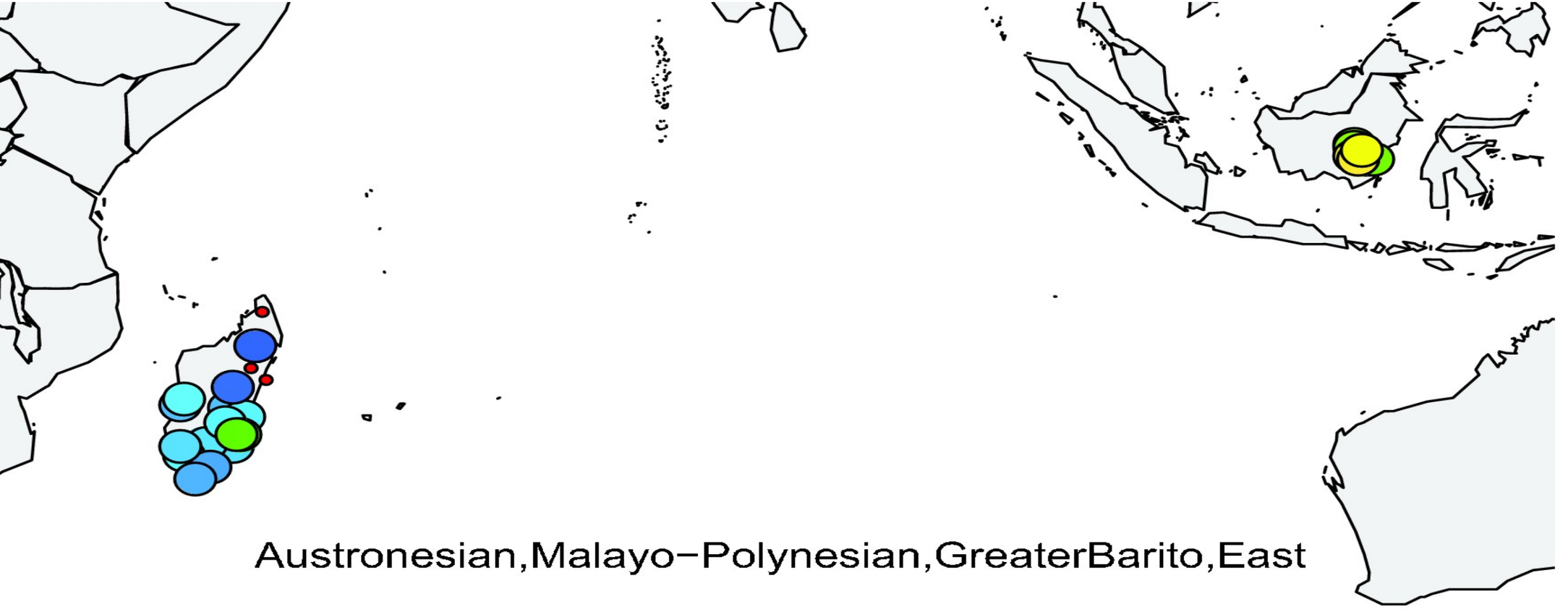


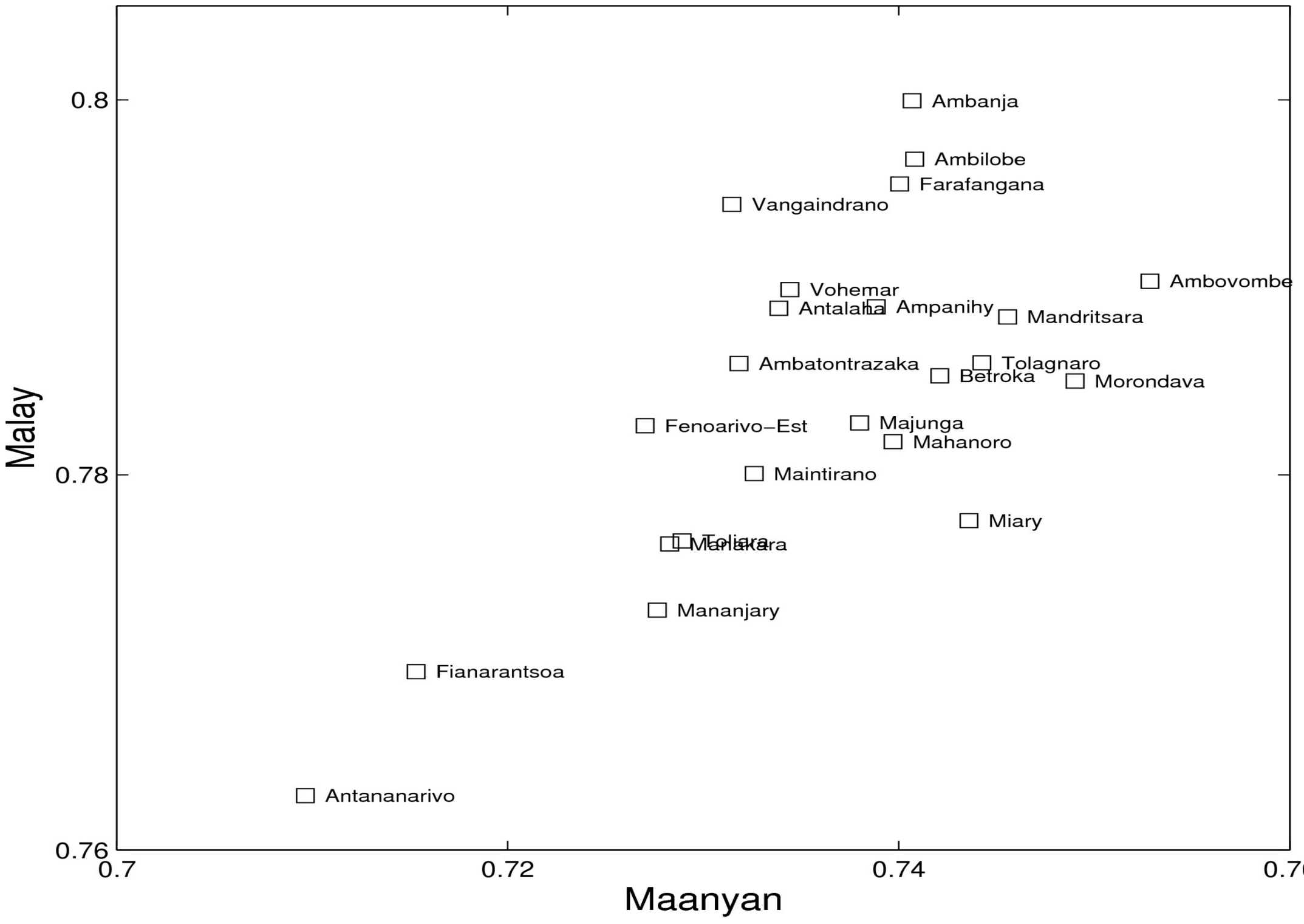
- The homeland of Malagasy dialects can be determined looking at the geographical area of maximal language diversity in the Island (Wichmann *et al.*).
- Furthermore, it can be assumed that the homeland corresponds to the landing place.
- The brownish-colored towns have the highest diversity values. The best candidate for the homeland is Manakara, but the surrounding towns are almost as good candidates. The northern locations are the less diverse and should have been settled last.
-



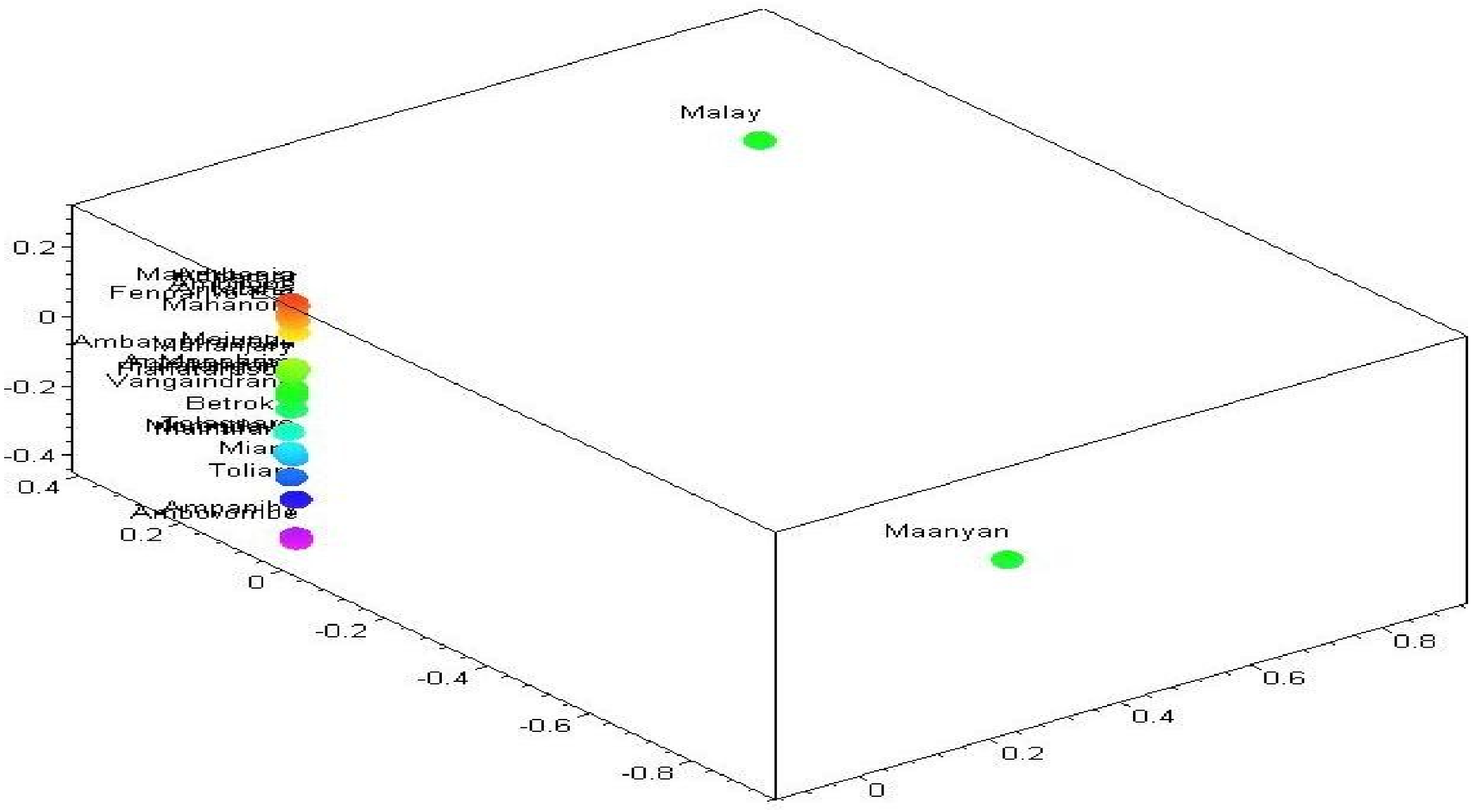
MALAGASY

- We now show the (still poor) results we have comparing Malagasy dialects with other Malayo-Polynesian languages.
- In next slide we consider Malagasy together with other languages of the Greater Barito East group. The homeland of Malagasy is in the same place but it is now secondary with respect to the Borneo homeland of the group.
- Then, we compute the distance of each of the Malagasy dialects from Maanyan and Malay and we show them on the associated Cartesian plane. Malagasy dialects seem to have the same relative composition (it may be a little less Malay and more Maanyan in the North, the contrary in the South) but some of them changed less (Antananarivo, Fianarantsoa, Manajary, Manakara) and some of them more (Ambovombe, Ambilobe, Ambanja).


Austronesian,Malayo−Polynesian,GreaterBarito,East


Malay
Maanyan
0.8
0.78
0.76
0.7
0.72
0.74
0.76
Ambanja
Ambilobe
Farafangana
Vangaindrano
Ambovombe
Vohemar
Antalaha
Ampanihy
Mandritsara
Ambatontrazaka
Tolagnaro
Betroka
Morondava
Fenoarivo−Est
Majunga
Mahanoro
Maintirano
Miary
Toliara
Morafeno
Mananjary
Fianarantsoa
Antananarivo

- This picture, obtained by Structural Component Analysis, confirms the findings of the previous slide. Colors individuate the plane, Malagasy dialects are orthogonal to the plane containing also Maanyan and Malay but some of them (green) are closer.



Conclusions

- RESULTS:
- A new internal classification of dialects,
- A date and a place for landing,
- Evidence that some central and south-east dialects changed less and they are closer to Malayo-Polynesian languages.

- OPEN PROBLEMS:
- The mystery concerning ancestry: it would be interesting to isolate and to quantify the different Malayo-Polynesian contributions to Malagasy.

REFERENCES

M. Serva and F. Petroni, *Indo-European languages tree by Levenshtein distance*. EuroPhysics Letters 81, (2008), 68005.

E. W. Holman, S. Wichmann, C. H. Brown, V. Velupillai, A. Müller and D. Bakker, *Explorations in automated lexicostatistics*. Folia Linguistica 42.2, (2008), 331-354.

F. Petroni and M. Serva, *Languages distance and tree reconstruction*. Journal of Statistical Mechanics: theory and experiment, (2008), P08012.

D. Bakker, A. Muller, V. Vellupillai, S. Wichmann, C. H. Brown, P. Brown, D. Egerov, R. Mailhammer, A. Grant and E. W. Holman, *Adding typology to lexicostatistics: a combined approach to language classification*, Linguistic Typology 13, (2009), 167-179.

P. Vérin, C. P. Kottak and P. Gorlin, *The glottochronology of Malagasy speech communities*, Oceanic Linguistics 8, (1969), 26-83

Ph. Blanchard, F. Petroni, M. Serva and D. Volchenkov. *Geometric Representations of Language Taxonomies*. Computer Speech and Language, (in press).

Søren Wichmann, André Müller and Viveka Velupillai. *Homelands of the world's language families*, Diachronica 27:2 (2010), 247–276.