SPATIAL DISTRIBUTION OF THE DIFFERENT STRAINS OF THE DISTINCT COCONUT LETHAL YELLOWING TYPE PHYTOPLASMA SPECIES ASSOCIATED WITH THE SYNDROME IN TANZANIA.

Anatolia Mpunami[1], Fabian Pilet[2,3], Sandrine Fabre[2,4], Allois Kullaya[1], Matthew Dickinson[5], Michel Dollet[2,6]

[1]Mikocheni Agricultural Research Institute, P.O. Box 6226, Dar-es-Salaam, Tanzania

[2]UMR IPME, CIRAD, 34398 Montpellier, France

[3]Present address: UMR PVBMT, CIRAD, 97410, Saint-Pierre, La Réunion, France

[4]Present address: UMR BGPI, CIRAD, 34398 Montpellier, France

[5]Sutton Bonington Campus, University of Nottingham, LE12 5RD Loughborough, UK

[6] Present address: Quarantine and plant health, Embrapa Recursos Genéticos e Biotecnologia, 70770-917, Brasilia DF, Brazil

Corresponding author: dolletm.palm@gmail.com; ORCID: 0000-0002-7793-3445

Abstract

Phytoplasmas are associated with the Lethal Disease of coconut palms in Tanzania (LDT). It is a destructive Lethal Yellowing Type Syndrome (LYTS) exhibiting differences in losses between the southern districts and the northern districts. To explain these differences, the existence of variable pathogenic strains of the LDT phytoplasma was investigated using ribosomal RNA gene PCR primers. A total of 84 samples were collected from 67 palms in 14 coastal districts of Tanzania, including the low, moderate and high incidence areas. Of these, 38 samples were studied in detail. Detected phytoplasma rDNA was characterized by either sequencing of the PCR products and/or restriction fragment length polymorphism (RFLP).

Sequence analysis of the P1/P7-primed PCR products revealed several positions of variability, making it possible to distinguish two main geographical clusters. The northern cluster included samples from Tanga region only and is associated with low/moderate disease incidence. A second larger cluster included samples from the rest of the coastline between Bagamoyo and Mtwara. Five genotypes could be identified based on mutations/deletions in the P1/P7 PCR product, two within the northern cluster, and three within the southern cluster. The geographical distribution of the two clusters and the genotypes could be related to the history of coconut introductions in Tanzania. The sequences obtained also confirm that the phytoplasmas associated with LDT are significantly different from all of the other phytoplasmas associated with coconut lethal yellowing type syndromes worldwide, and it is proposed that these phytoplasmas should be classified into their own 16Sr group.

Keywords: *Candidatus* phytoplasma cocostanzaniae. East African Tall coconut populations. Lethal Yellowing Type Syndrome. Phylogenetic tree. RFLP. rRNA gene.

Introduction

Coconut palm (*Cocos nucifera* L.) is the most important perennial oil crop that supports the livelihood of most farmers in the coastal areas of Tanzania. Currently, 25 million palms are grown on 262,000 ha along the coastal belt of mainland Tanzania and the islands of Zanzibar, Pemba and Mafia, where more than 95% of palms are cultivated by smallholder farmers (Kullaya and Mpunami 2008). The crop is however, threatened by a destructive Lethal Yellowing-Type Syndrome (LYTS) known as "Lethal Disease" or "Lethal Disease Tanzania" (LD or LDT) or even "Tanzanian lethal decline" (Schuiling et al. 1992a) (Figure 1). The symptoms were described before (Steiner 1978; Schuiling et al. 1981). Since 1965, LDT has killed more than nine million palms or 40% of the groves (Schuiling et al. 1992b).

Phytoplasma are consistently associated with the disease (Schuiling et al. 1981, Nienhaus et al. 1982; Mpunami et al. 1999). LDT was first reported in the country in 1905 near Bagamoyo, then in 1912 in the south close to the Mozambique border, and at Kunduchi, close to Dar-es-Salaam (Schuiling et al. 1992b). However, the most serious outbreaks occurred independently in the 1940s in the districts of Kilwa, Rufiji (south of the Pwani region) (Figure 2) and in the 2000s, in Mkuranga (centre of Pwani).

There exists a profound difference in the epidemiology of the LDT between affected areas. In the northern districts of Tanga region and Bagamoyo district, the incidence of disease was low to moderate in the early 2000s (average losses of 0.6% per year). However, in the southern coastal districts of Kilwa, Rufiji and Mkuranga, the disease was very severe and losses to LDT were over four times higher than in the northern districts (Schuiling et al. 1992b, Mpunami et al. 2008). These differences have been presumed to be related to the diversity in the coconut germplasm planted in these different regions (Schuiling et al. 1992b).

PCR using primer pairs based on the 16S rRNA gene, then on the 16-23S intergenic spacer region have been previously developed and used extensively for detection of LDT phytoplasmas in both plant and insect host tissues (Rohde et al. 1993; Tymon et al. 1998; Mpunami et al. 1999; 2000).

Molecular studies based on the ribosomal operon have demonstrated that there is an important diversity among the phytoplasma associated with the various LYTS in the world (Harrison et al. 1994; Tymon et al. 1998; Harrison et al. 2002; Harrison et al. 2008; Marinho et al. 2008; Martinez et al. 2008; Dollet et al. 2009; Harrison et al. 2014; Pilet et al. 2019).

Harrison et al. (2014) noted that the 16SrIV phytoplasmas of the Americas should be classified as '*Candidatus* Phytoplasma palmae' whilst the 16SrXXII-A type of Nigeria and Mozambique, should be classified as '*Ca.* Phytoplasma palmicola' and the 16SrXXII-B type found in Ghana and Côte d'Ivoire as '*Ca.* Phytoplasma palmicola'-related. They also noted that the LDT phytoplasma from Tanzania, which has historically been classified as 16SrIV-C, is distinct from both of the other two subclades, and it had been informally proposed that these subclades represented three separate candidate species of '*Ca.* Phytoplasma'. In the same article, a name of "provisional candidate species" for the phytoplasma associated with the LYTS in Tanzania was proposed: '*Ca.* Phytoplasma cocostanzaniae'. However, despite the differences in sequences and epidemiology, these three distinct subclades are often mistakenly clumped together and referred to as lethal yellowing disease (Originally, "Lethal Yellowing" is the name used for the disease in the Americas) rather than distinct "Lethal Yellowing Type Syndromes" (LYTS).

In the present study, phytoplasmas associated with 67 LDT-diseased coconut palms in 14 districts throughout coastal Tanzania were analyzed in order to determine whether or not differences in disease incidence was associated with genetic variability among phytoplasma populations. We report detection of five different genotypes/strains differently distributed according to the variability of the spread of the disease. In addition, the16Sr RNA gene sequences obtained have been used to re-emphasize that the LDT phytoplasmas are different from those associated with LY of the Americas (16SrIV) and from '*Ca.* Phytoplasma palmicola' (16SrXXII) from West Africa and Mozambique.

Materials and Methods

Plant material

Eighty-four samples of immature inflorescences and/or leaflets of spear leaves from 67 East African Tall (EAT) coconut palms displaying typical LDT symptoms were collected. Palms for sampling were selected from the farmers' fields in different villages within the 14 coastal districts that were affected by LDT. The districts sampled include Mkinga and Tanga from the north (low incidence areas); Pangani, Bagamoyo, Kinondoni Temeke and Mkuranga from north and central Tanzania (moderate incidence), Rufiji, Kilwa and Lindi in the south (high incidence areas) (Figure 2). The collection also included districts of Mtwara, Tandahimba and Newala, further south in an area that had previously low incidence but in which the incidence of disease has recently increased in the early 2000s. Only samples of the diseased trees showing a strong electrophoresis band exploitable for direct sequencing or RFLP are listed (Table 1).

Leaflets were excised from the immature spear leaves, and immediately placed in plastic bags for transportation to the laboratory at Mikocheni Agricultural Research Institute (MARI) in Dar-es-Salaam. For inflorescences, whole unopened inflorescences were excised from affected palms and transported to the laboratory. On arrival, the inflorescences were split open, and flowers removed from the rachillas. The rachillas were cut off and crushed for DNA extraction, or frozen at – 20 °C until DNA was extracted. Similarly, the leaflets were immediately crushed for DNA extraction, or frozen at – 20 °C.

DNA extraction

DNA was extracted from the spear leaves or rachillas, using a modification of the CTAB extraction protocol of Doyle and Doyle (1990). Five grams of freshly harvested or frozen tissues were used. Nucleic acids pellets were resuspended in 1ml of 1 x TE buffer, pH8·0.

PCR Analysis

Two different primer combinations were used to amplify phytoplasma rDNA from LDT affected palm tissues. First, amplification of part of the 16S rRNA gene was performed using the primer pair Rohde forward and Rohde reverse (Rohde et al. 1993). The second primer pair was the phytoplasma universal primer pair P1 / P7 (Deng and Hiruki 1991; Smart et al. 1996) which covered most of the 16S rRNA gene, the 16S-23S rRNA intergenic region and the 5'end of the 23S rRNA gene (Figure 3).

For each PCR, a 25 µL reaction mixture contained about 50 ng template DNA, 150 mM mixed deoxynucleotide triphosphates (dNTP), 50 ng of each primer, 1 Unit of *Taq* Polymerase, and PCR buffer 1X (PCR Master Kit, Qiagen). For the Rohde's primer pair, the mixture was subjected to 36 cycles (Gene Amp PCR System 9700, Applied Biosystems) with the following parameters: 94°C for 60 s (after a first denaturation of 60s), 57°C for 80 s, 72°C for 130 s, followed by a final extension of 72°C for 5 min.

The PCR parameters for the P1 /P7 primer pair were: 94°C for 90 s followed by 35 cycles of 94°C for 30 s, 56°C for 50 s, 72°C for 90 s and a final extension of 72°C for 10 min. For positive control, a sample of LDT DNA that was previously confirmed at CIRAD, or clover phyllody positive DNA, was used.

PCR products were analyzed by electrophoresis through a 1% agarose gel and visualized in the gel by UV transillumination after staining with ethidium bromide.

DNA Sequencing and Sequence analysis

Twelve PCR products using Rohde's primers and 15 samples amplified using P1/P7 were sequenced (Beckman Coulter Genomics). Rohde's PCR products were sequenced using Rohde forward primer only. For the P1/P7 PCR, the two strands were sequenced using both P1 and P7. Accession numbers of the sequenced products are listed in Table 1.

Sequences were edited and aligned along with those of LDT phytoplasma accessions X80117 and EU168773, and sequences of phytoplasmas from other 16Sr groups obtained from GenBank. Phylogenetic analysis was performed in Mega v. 6.06 (Tamura et al. 2013) using the package ClustalW followed by a phylogenetic reconstruction by neighbour-joining using the bootstrap method (with 1000 replications) as a test of phylogeny and maximum composite likelihood as the model.

Based on the variable regions identified in the aligned sequences, putative endonuclease recognition sites and the relevant restriction enzymes that were likely to show differences between LDT samples were identified using the Vector NTI software (Invitrogen).

RFLP Analysis of PCR Products

Variability within the 16S rRNA gene amplified by primer pair P1/P7 was identified by digestion of the P1/P7 product with *BstU*I endonuclease, which was the unique enzyme among the 17 commonly used for phytoplasmas (Lee et al. 1998) able to differentiate our samples. Six microliters (6 µL) of the P1/P7 PCR product (for selected samples) was digested with 5 U of *BstU*I, (New England Biolabs) at 60 °C overnight in 10 µL reaction volumes. Digests were separated on 2% agarose gels in 1 X TAE buffer, stained with ethidium bromide, and visualized under UV light.

In silico digestions of the F2nR2 region with 17 restriction enzymes were performed using pDRAW32 software (AcaClone Software, http://www.acaclone.com) as previously described (Wei et al., 2007; 2008), Similarity coefficients were derived from the relationship $F=2N_{xy}/(N_x+N_y)$ (Nei et al. 1979) in which x and y are the two given strains under investigation, N_x and N_y are the total fragments resulting from digestion by 17 enzymes in strain x and y respectively, and N_{xy} is the number of fragments shared by the two strains.

Results

Detection of LDT phytoplasma using the primer pair Rohde forward/Rohde reverse

The Rohde forward/Rohde reverse primer pair, designed to detect the LDT phytoplasma, was used to analyze the DNA extracted from all 84 collected palm samples. A 450 bp band was resolved by agarose gel electrophoresis from the positive control LDT DNA and from all 84 samples, except three. One of the three palms that did not amplify a product with this LDT-specific primer pair was probably suffering from drought rather than LDT.

Detection of LDT phytoplasma using the phytoplasma universal primer pair P1/ P7

In order to study a larger fragment of the phytoplasma 16S rDNA gene and the 16S-23S rDNA intergenic region, DNA from 38 samples representative of the low, medium and high disease incidence areas were amplified using the phytoplasma universal primer pair P1/P7. A 1.8 Kb PCR product was amplified from all 38 samples and from the clover phyllody DNA (positive control), but not from the water control (Figure 4).

Analysis of genetic variation between isolates by sequencing

Amplification products obtained with the Rohde primer pair and with P1/P7 were both used to search for genetic variation between the different isolates.

Products obtained with Rohde's primers from 12 EAT coconuts were sequenced. The alignment of the 12 sequences did not reveal any variability throughout the entire 500 bp of the sequenced regions, excepted for the isolate Tanz 08-06, which showed one mutation (R1) (Figure 3) at the position 1002 (Table 2). First position has been arbitrarily defined as the first base of the P1/P7 theoretical product ($A^1A^2G^3A^4G^5T^6T^7T^8G^9A^{10}…$).

8

Further analysis was done on a selection of PCR products amplified from 17 geographically representative isolates using the primer pair P1/P7. Sequence alignment revealed a number of variations/mutations/deletions which could group LDT samples into 5 different genotypes named TZ-I to TZ-V (Table 2). The first genotype (TZ-I) differs from the second one (TZ-II) characterized by a double mutation (M1:451; M2:460) inside the 16S rDNA corresponding to dimers in the secondary structure of the 16S rRNA (Table 2). A third mutation (M3:1242) was found inside the 16S rDNA and defined the genotype TZ-III. This mutation occurred in the middle of the sequence of the Rohde reverse primer. No mutation was observed inside the Ile tDNA. Two deletions were found. The first one (M4:1583) was located inside the internal transcribed spacer region (ITS1) and defined the TZ-IV genotype, whereas the second one (M5:1698-1699) corresponds to a double deletion localized in the ITS2, and separated out genotype TZ-V.

The sequence [X80117], which was derived from a sample collected in a 1993 survey from the Chambezi/Kifumangao region close to Bagamoyo, covering a large part of the 16S rDNA gene, showed 100% of homology with the samples of the genotype TZ-III, IV and V whereas the sequence [EU168773] (Hodgetts et al. 2008), also derived from a sample from the same 1993 survey, but which only covered the 16S-23S ISR, matched with the genotype TZ-I to TZ-III at 100% (Table 2).

One last mutation could be observed at the position 1759. However, this mutation was not considered because of the low number of sequences workable at this position.

Although grouped as genotype TZ-III, the isolate Tanz 08-78, is considered a variant due to the same mutation (R1) as observed for the isolate Tanz 08-06 on the Rohde primer sequences. According to this classification, the samples with Accession X80117 and EU168773 probably belong to the genotype TZ-III, and indicate that this sequence type was present in the central region in the early 1990s.

One isolate (Tanz 08-18), presented ambiguous results. The sequence chromatogram displayed the mutations M1 and M2, but for each of those positions, a double signal of equal intensity was shown. At the position 457 (M1), both G and A are present whereas both T and C can be observed at position 460 (M2).

Because the two sequences of the sample Tanz 08-74 were not contiguous, we were not able to assemble them in one unique sequence. The two sequences have been submitted for Accession Numbers independently. However, according to the mutations observed, this sample has been grouped as genotype TZ-III, the 56 bp missing corresponding to an unvariable sequence.

Mapping of the different genotypes revealed a gradient from the north to the south of Tanzania (Figure 2). The genotypes TZ-I and TZ-II were present in the northern districts of Mkinga, Tanga and Pangani only. The genotype TZ-III was observed in the Pwani (Coast) region and in the north of the Lindi region (Kilwa district). The genotype TZ-IV was present in the center and the south of the Lindi region (district of Kilwa and Lindi). The last genotype TZ-V has been observed exclusively in the Mtwara region.

Analysis of variation between isolates by restriction fragment length polymorphisms

To be able to evaluate more LDT isolates, the enzyme *BstU*I was used to digest the PCR products amplified from 38 LDT samples with P1/P7.

Two patterns of restriction fragments were obtained (Figure 5). The first pattern (RFLP-A) is characterized by one band of 816 bp whereas the pattern RFLP-B displays one band of 728 bp. The sizes of the three following bands are too close to be differentiated on agarose gel. The last band of 88 bp can be observed only for the pattern RFLP-B.

Of the 38 digested samples, 10 samples belong to the group RFLP-A (Table 1) and are categorized as either the genotype TZ-I or TZ-II. Those 10 samples originated from the north

of the country and correspond to the low to moderate disease incidence areas (Figure 2). The remaining 28 samples fall in the RFLP-B pattern and correspond to the genotypes TZ-III, TZ-IV and TZ-V. They originated from the regions of Pwani, Lindi and Mtwara (Figure 2). Virtual RFLP pattern with 17 restriction enzymes revealed only two different patterns with the *Bst*UI enzyme. A total of 45 bands were generated for the isolates of the RFLP-B group, and 46 for the RFLP-A group, with 44 bands being common to both groups, and giving a similarity coefficient F=0.967.

Phylogenetic relationship of LDT to other coconut Lethal Yellowing Type Syndrome phytoplasmas

To confirm the relationship of the current LDT strains to historic strains of LDT and to the LYTS phytoplasmas from the Americas and Africa, a phylogenetic tree was constructed using the 16S rRNA sequences for the 15 samples from this study for which P1/P7 sequences were obtained (Table 1), with sequences retrieved from GenBank. The phylogenetic tree (Figure 6), confirms the position of the LDT subclade as being distinct from the 16SrIV subgroups A, B, D, E and F subclade of the Americas, and from the '*Ca*. Phytoplasma palmicola' 16SrXXII subgroups A and B subclade from West Africa and Mozambique.

Discussion

This study is the first to show a regional distribution of different LDT phytoplasma genotypes. We demonstrated that the different LDT phytoplasma genotypes were not randomly distributed, but rather localized in spatially distributed ecological niches. The results, therefore, confirmed the hypothesis about local adaptation of the LDT pathogen to existing environmental conditions including the host population. The existence of different strains of

11

the LDT phytoplasma was one hypothesis to explain observed differences in disease incidence between various regions of Tanzania. However, genotype differences were not evident using Rohde's primers (Mpunami et al. 1999). For the first time, we report here a strong correlation between the different genotypes of the LDT phytoplasma and the different levels of disease incidence in Tanzania.

Sequence analysis of the 450 bp rDNA sequences from 12 representative LDT samples with Rohde's primers revealed no variability and confirm the previous observation of Mpunami et al. (1999) of nine samples from different areas amplified with these primers. Only one sample, from Kilwa district, displayed a single base substitution. However, using the P1/P7-primed 1.8 Kbp PCR products from 15 representative LDT samples in this current study, the LDT samples could be divided into five phytoplasma genotypes.

This study has revealed one mutation (M3) inside the Rohde's reverse primer. This mutation may have previously affected PCR assays and resulted in false negatives. Moreover, these primers are not able to identify the two subgroups present in Tanzania. For this, it would be necessary to define a new primer pair enclosing the M3 mutation to generate a large PCR product which can be used for restriction enzyme digestion analysis.

Direct sequencing of PCR products without the selective cloning step revealed two sequences for one of our samples (Tanz 08-18). This result could be due to the presence of two heterogenic rRNA operons in the strain as previously described for several phytoplasma strains (Harrison et al. 2004; Liefting et al. 1996; Jomantiene et al. 2002). However, the single 16S-23S sequence for the other samples and the presence of the genotype TZ-I and TZ-II in the Tanga region (Figure 2) may also suggest a mixed infection.

Two genotype clusters were identified using *Bst*UI restriction enzyme, the first (RFLP A) grouping two genotypes, TZ-I and TZ-II, and the second (RFLP B) grouping the genotypes TZ-III to TZ-V. Samples of the subgroup A are located in the north of Tanzania and are associated with low to moderate disease incidence. According to Schuiling and Harries (1994), the Tanga region is an area where coconuts were first introduced in Tanzania several centuries ago, probably from India. Furthermore, experience gained from years of field experimentation with coconut materials collected from these districts and other parts of the country has demonstrated that some populations of East African Tall palms from Tanga region are more tolerant to LDT even when planted in areas of high disease incidence (Schuiling et al. 1992b, Kullaya et al. 1995, Mpunami et al. 2008). This suggests that some introductions of the EAT coconut in the area possess disease resistance genes, considering that the introductions were gradually done over an extended period of time and from different sources, hence the heterogeneous nature.

The B subgroup samples occur in regions in which the incidence of disease ranges from moderate to very high. In most of these regions, coconuts were introduced fairly recently when compared to Tanga region. In the area covered by the subgroup, it is possible to distinguish three regions with three different levels of disease incidence. In the Bagamoyo district (North of the subgroup), where the LDT was first diagnosed in 1905, the coconut industry is comparatively older, and the incidence of disease is moderate. Whilst no P1/P7 sequences were derived from samples from this region in the current study, the historic sample that was used to generate the LDT sequences in the GenBank database of X80117 and EU168773 came from this area, and appears to be of genotype TZ-III. The samples from the area between Bagamoyo in the Coastal region and Kilwa district in Lindi region have this

same genotype TZ- III but are associated with the very high incidence of disease, causing complete devastation during the past forty years. In the South, in the districts of Mkuranga, Rufiji and Kilwa, where coconut growing was intensified more recently during the German colonial era (1885-1920) using uniform planting material from Mafia Island, extensive areas have been devastated by the disease. The genotype TZ-V from the southern regions of Lindi and Mtwara correspond to areas where rapid disease spread is a recent phenomenon. Although LDT was first reported in the area of Mtwara nearly a century ago (1912), the disease did not spread in these districts until 1998 when it was first observed to be spreading rapidly.

Interestingly, there is a close correlation between the distribution of the LDT phytoplasma strains identified and the distribution of the EAT coconut genotypes found along the coast of Tanzania. In a study conducted by Kullaya et al. (2001), two main clusters of palm populations were revealed, with one cluster of sub-populations from Tanga region only (northern part of the coastal belt), and another cluster composed of sub-populations from Bagamoyo (central part) and the southern districts of Lindi and Mtwara regions. The studies thus confirmed that the EAT is genetically a very heterogeneous population, even if globally, the coconuts in Tanzania result from the introgression of germplasm from South-East Asia into a genetic background that originated from South Asia (Gunn et al. 2009).

In addition to the data on the separation of the LDT population into the A and B subgroups in this study, we have also re-analysed the relationship of these phytoplasmas to the others that have been reported as being associated with the LYTS. In the Americas, phytoplasmas of group 16SrIV occur in coconut and several other palm species. In this group it is possible to distinguish at least 5 subgroups, -A, B, D, E and F, and the LY disease of coconut is mainly associated with subgroup 16SrIV-A. The phytoplasmas of subgroup 16SrIV-D have also been

detected in coconut but are mainly detected in other palms such as *Phoenix* spp., in which they cause a decline called Texas Phoenix palm decline (TPPD). The LDT phytoplasma, which was previously classified as 16SrIV-C within this system, is clearly in a distinct subclade (Figure 5) as it has been suggested before (Tymon et al. 1998, Mpunami et al. 1999; Marinho et al. 2008; Harrison et al. 2014). We advocate that the 16SrIV-C should no longer be used for this phytoplasma and it should be given a new 16Sr group designation. The candidate species could be the one already suggested '*Ca.* Phytoplasma cocostanzaniae'. However, as the same phytoplasma occurs in diseased coconuts in Kenya and in the north of Mozambique, 'cocostanzania' may be too restrictive. The 16SrXXII-A '*Ca.* Phytoplasma palmicola' and 16SrXXII-B '*Ca.* Phytoplasma palmicola'-related strains from west Africa and Mozambique clearly form a third distinct subclade, as previously reported by Harrison et al. (2014).

Additional phylogeny studies on both the LDT phytoplasma and the coconut host from Tanzania, Kenya and Mozambique using more polymorphic genes and microsatellites to increase precision and involving more samples could help to understand the relationship between the phytoplasmas, the different level of resistance or tolerance to LDT and their co-evolution in East Africa.

Author contribution Statement

AM: Field survey, investigation, review & editing

FP: Investigation, validation, visualization, co-writing – original draft, review & editing

SF: Resources, investigation, validation

AK: Field survey, project administration (Tanzania), resources

Matthew D: Investigation, supervision, validation, visualization, review & editing

Michel D: Conceptualization, project administration (France), funding acquisition, resources, investigation, supervision, validation, visualization, co-writing – original draft, review & editing

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Dollet M, Quaicoe RN, Pilet F (2009) Review of Coconut "Lethal Yellowing" type diseases Diversity, variability and diagnosis. OCL 16:97-101

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13-5

Deng S, Hiruki C (1991) Amplification of 16S rRNA genes from culturable and nonculturable mollicutes. Journal of Microbiological Methods 14:53-61

Gunn BF, Baudouin L, Olsen KM (2009) Independent origins of cultivated coconuts (*Cocos nucifera* L.) in the Old-World Tropics. PLoS ONE 6(6): e21143 doi:10.1371/journal.pone

Harrison NA, Richardson PA, Jones P, Tymon A, Eden-Green SJ, Mpunami A (1994) Comparative investigation of mycoplasma-like organisms (MLOs) associated with Caribbean and African coconut lethal decline diseases by DNA dot hybridization and PCR assays. Plant Disease 78:507-511

Harrison NA, Myrie W, Jones P, Carpio ML, Castillo M, Doyle MM, Oropeza C (2002) 16S rRNA interoperon sequence heterogeneity distinguishes strain populations of palm lethal yellowing phytoplasma in the Caribbean region. Annals of Applied Biology 141:183-93

Harrison NA, Carpio ML, Davis RE, Jomantiene R (2004) Organization of two heterologous rRNA operons and adjacent genes in the palm lethal yellowing phytoplasma genome. In: Proceedings of the Fifteenth International Congress of the International Organization for Mycoplasmology; 2004. Athens, Georgia. p.127

Harrison NA, Helmick EE, Elliott ML (2008) Lethal yellowing-type diseases of palms associated with phytoplasmas newly identified in Florida, USA. Annals of Applied Biology 153:85-94

Harrison N, Davis RE, Oropeza C, Helmick E, Narvaez M, Eden-Green S, Dollet M, Dickinson M (2014) '*Candidatus* Phytoplasma palmicola', a novel taxon associated with a lethal yellowing-type disease (LYD) of coconut (*Cocos nucifera* L.) in Mozambique. Journal of Systematic and Evolutionary Microbiology 64:1890-1899. DOI 10.1099/ijs.0.060053-0

Hodgetts J, Boonham N, Mumford R, Harrison N, Dickinson M (2008) Phytoplasma phylogenetics based on analysis of *secA* and 23S rRNA gene sequences for improved resolution of candidate species of '*Candidatus* Phytoplasma'. International Journal of Systemic and Evolutionary Microbiology 58**:**1826-1837

Jomantiene R, Davis RE, Valiunas D, Alminaite A (2002) New group 16SrIII phytoplasma lineages in Lithuania exhibit rRNA interoperon sequence heterogeneity. European Journal of Plant Pathology 108**:** 507-517

Kullaya A, Mpunami A, Harries HC, Schuiling M, Kaiza DA (1995) Screening for resistance to the coconut Lethal Disease in Tanzania and prospects for rehabilitation. In: Eden-Green SJ, Ofori F (Eds.) Proceedings of the International Workshop on Lethal Yellowing-Like Diseases of Coconut, Elmina, Ghana. pp.168-72

Kullaya AK, Mpunami A (2008) Breeding for resistance to coconut Lethal Disease in Tanzania. In: Proceedings of an International Workshop on Lethal Yellowing Disease of Coconut, 3–5 June 2008, Accra, Ghana. CIRAD (Centre de coopération internationale en recherche agronomique pour le développement), France

Kullaya A, Mkumbo K, Sallu R, Tembo S, Masumbuko L (2001) Characterization of the local coconut Germplasm using RAPD markers. Breeding and Seed Production. In: Twenty-five Years of Coconut Research in Tanzania. Mikocheni Agricultural Research Institute.

Lee IM, Gundersen-Rindal DE, Davis RE, Bartoszyk IM (1998) Revised classification scheme of phytoplasmas based on RFLP analyses of 16S rRNA and ribosomal protein gene sequences. International Journal of Systematic Bacteriology 48:1153-1169

Liefting LW, Andersen MT, Beever RE, Gardner RC and Forster RLS (1996) Sequence heterogeneity in the two 16S rRNA genes of *Phormium* yellow leaf phytoplasma. Applied and Environmental Microbiology 62: 3133-3139

Marinho VL, Fabre S, Dollet M (2008) Genetic variability among isolates of Coconut lethal yellowing phytoplasmas determined by Heteroduplex Mobility Assay (HMA). Tropical Plant Pathology 33: 377-380

Martinez RT, Narvaez M, Fabre S, Harrison N, Oropeza C, Dollet M, Hichez E (2008) Coconut lethal yellowing on the southern coast of the Dominican Republic is associated with a new 16SrIV group phytoplasma. Plant pathology 57**:**366

Mpunami AA, Mugini JA, Tembo PM, Chalamila S, Luka FS, 2008. Epidemiology of the Coconut Lethal Disease in Tanzania In: *Proceedings of the International Workshop on Lethal Yellowing Diseases on Coconut*, 2008. Accra, Ghana, 170-82

Mpunami AA, Tymon AM, Jones P, Dickinson MJ (1999) Genetic diversity in the coconut lethal yellowing disease phytoplasmas of East Africa. Plant Pathology 48:109-114

Mpunami A, Tymon A, Jones P, Dickinson MJ (2000) Identification of potential vectors of the coconut lethal disease phytoplasma. Plant Pathology 49:355-361

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the USA 76:5269–5273. doi: 10.1073/pnas.76.10.5269

Nienhaus F, Schuiling M, Gliem G, Schninzer V, Spittel A (1982) Investigations on the etiology of lethal disease of coconut palm in Tanzania. Zeitschrift fur Pflanzenkranheiten und Pflanzenschutz 89:185-193

Pilet F, Quaicoe RN, Osagie IJ, Freire M, Foissac X (2019) Multilocus sequence analysis reveals three distinct populations of "*Candidatus* Phytoplasma palmicola" with a specific geographical distribution on the African continent. Applied and Environmental Microbiology, 85:e02716-18. doi:10.1128/AEM.02716-18

Rohde W, Kullaya A, Mpunami A, Becker D (1993) Rapid and sensitive diagnosis of mycoplasma-like organisms associated with lethal disease of coconut palm by a specially primed polymerase chain reaction for the amplification of 16S rDNA. Oleagineux 47:511-515

Schuiling M, Nienhaus F, Kaiza DA (1981) The syndrome in coconut palms affected by a lethal disease in Tanzania. Zeitschrift fur Pflanzenkranheiten und Pflanzenschutz 88:665-677

Schuiling M, Mpunami A (1992) Lethal disease of coconut palm in Tanzania. I. Comparison with other coconut diseases in East Africa Oleagineux 47:511-515

Schuiling M, Mpunami A, Kaiza DA (1992) Lethal disease of coconut palm in Tanzania. II. History, distribution and epidemiology Oléagineux 47:516-522

Schuiling M, Harries HC (1994) The coconut palm in East Africa. 1. East African Tall. Principes 38:4-11

Smart CD, Schneider B, Blomquist CL, Guerra LJ, Harrison NA, Ahrens U, Lorenz KH, Seemuller E, Kirkpatrick BC (1996) Phytoplasma-specific PCR primers based on sequences of the 16S-23S rRNA spacer region. Applied and Environmental Microbiology 62: 2988-2993

Steiner KG (1978) Suspected lethal yellowing disease of coconut palms in Tanzania. FAO Plant Protection Bulletin 26:10-12

Tamura K, Stecher G, Paterson D, Filipski A, Kumar S (2013) Molecular evolutionary genetics analysis version 6.0. Molecular Biology and
 Evolution 30:2725–2729

Tymon A, Jones P, Harrison NA (1998) Phylogenetic relationships of coconut phytoplasmas and the development of specific oligonucleotide PCR primers. Annals of Applied Biology 132:437-452

Wei W, Davis RE, Lee IM, Zhao Y (2007) Computer-simulated RFLP analysis of 16S rRNA genes: identification of ten new phytoplasma groups. International Journal of Systematic and Evolutionary Microbiology 57:1855-1867

Wei W, Lee IM, Davis RE, Suo X, Zhao Y (2008) Automated RFLP pattern comparison and similarity coefficient calculation for rapid delineation of new and distinct phytoplasma 16Sr subgroup lineages. International Journal of Systematic and Evolutionary Microbiology 58, 2368-2677

Table 1 Variability among the sequences of Tanzanian coconut samples studied using primer pairs P1/P7 or Rhode's primers and the corresponding identified genotypes and strains/clusters: geographic origin, GeneBank accession numbers, genotype group and RFLP group.

| Sample Numbers	Organs	Region	District	Location	GenBank accession no. Rhode	GenBank accession no. P1P7	Genotype	RFLP Group
Tanz 08-01	Inflorescence	Mtwara	Mtwara Rural	Mnawene	GU952082	GU952107	TZ- V	B
Tanz 08-02	Inflorescence	Lindi	Kilwa	Mirumba, Kilwa	GU952088	GU952102	TZ- IV	B
Tanz 08-03	Inflorescence	Mtwara	Mtwara Rural	Mbuo				B
Tanz 08-04	Inflorescence	Mtwara	Tandahimba	Huruma	GU952089		TZ- V	B
Tanz 08-05	Inflorescence	Mtwara	Mtwara Rural	Mpapura	GU952090	GU952106	TZ- V	B
Tanz 08-06	Inflorescence	Lindi	Kilwa	Mtoni	GU952080			B
Tanz 08-07	Inflorescence	Mtwara	Tandahimba	Namahonga				B
Tanz 08-08	Inflorescence	Lindi	Kilwa	Singino				B
Tanz 08-09	Inflorescence	Mtwara	Newala	Mnalale	GU952091			B
Tanz 08-10	Inflorescence	Lindi	Kilwa	Kikanda	GU952081	GU952099	TZ- III	B
Tanz 08-11	Inflorescence	Mtwara	Mtwara Rural	Madimba	GU952083	GU952104	TZ- V	B
Tanz 08-12	Inflorescence	Lindi	Kilwa	Ngwanga, Kinjumbi				B
Tanz 08-13	Inflorescence	Mtwara	Mtwara Rural	Nchepa				B
Tanz 08-14	Inflorescence	Lindi	Lindi Rural	Kitomanga	GU952084	GU952103	TZ- IV	B
Tanz 08-15	Inflorescence	Lindi	Lindi Rural	Lindi				B
Tanz 08-16	Inflorescence	Tanga	Tanga	Kisimatiu 1	GU952085	GU952094	TZ- I	A
Tanz 08-17	Inflorescence	Tanga	Tanga	Kisimatiu 2				A
Tanz 08-18	Inflorescence	Tanga	Mkinga	Mpirani, Gezani	GU952086	GU952092	TZ- I	A
Tanz 08-19	Inflorescence	Tanga	Pangani	Mnazi mmoja, Kimanga	GU952087	GU952093	TZ- I	A
Tanz 08-25	Leaves	Lindi	Kilwa	Kinjumbi				B
Tanz 08-32	Leaves	Lindi	Kilwa	Mzizima, Kilwa				B
Tanz 08-33	Leaves	Lindi	Kilwa	Miyumbu, Kinjumbi		GU952100	TZ- III	B
Tanz 08-39	Leaves	Lindi	Kilwa	Kilwa Masoko				B
Tanz 08-42	Leaves	Lindi	Kilwa	Kale, Kilwa Kinjumbi				B
Tanz 08-43	Leaves	Lindi	Kilwa	Masoko pwani, Kilwa		GU952101	TZ- III	B
Tanz 08-44	Leaves	Mtwara	Newala	Mkunya		GU952105	TZ- V	B
Tanz 08-50	Leaves	Pwani	Bagamoyo	Chambezi 2				B
Tanz 08-59	Leaves	Tanga	Pangani	Mikocheni, Mkwaja				A
Tanz 08-63	Leaves	Tanga	Tanga	Kibafuta Kikereni				A
Tanz 08-64	Leaves	Tanga	Mkinga	Boma				A
Tanz 08-65	Leaves	Tanga	Pangani	Boza				A
Tanz 08-67	Leaves	Dar es Salaam	Kinondoni	Mpiji				B
Tanz 08-69	Inflor.	Tanga	Tanga	Mabokweni		GU952095	TZ- II	A
Tanz 08-74	Leaves	Tanga	Mkinga	Vuo		GU952096-97	TZ- II	A
Tanz 08-75	Leaves	Pwani	Mkuranga	Binga				B
Tanz 08-77	Leaves	Pwani	Mkuranga	Dondo				B
Tanz 08-78	Leaves	Pwani	Mkuranga	Kisiju		GU952098	TZ- III	B
Tanz 08-84	Leaves	Pwani	Rufiji	Ruaruke				B

Table 2 Aligned sequences of 15 rDNA samples sequenced with P1/ P7 primers displaying mutations (**M**) that differentiate the five genotypes of LDT phytoplasma

| Isolate	Genotype	M1　　　　M2　　　　　　M3　　　　　　M4 M5
[X80117]		`AAAATAGTGGAAAAACTATATT//TGAAACGCGAGTTTTT`
[EU168773]		`//TTTTTTTGTAAGT//CTAATATATATGGCACCT`
Tanz 08-16	TZ-I	`AAAATAATGGAAAAATTATATT//TGAAACGTGAGTTTTT//TTTTTTTGTAA` `GT//CTAATATATATGGCACCT`
Tanz 08-19		`......A.......T......//......T........//......T....` `..//.........AT.......`
Tanz 08-69	TZ-II	`......G.......C......//......T........//......T....` `..//.........AT.......`
Tanz 08-74		`......G.......C......//......T........//......T....` `..//.........AT.......`
Tanz 08-18	TZ-I+II	`......R.......Y......//......T........//......T....` `..//.........AT.......`
Tanz 08-78	TZ-III	`......G.......C......//......C........//......T....` `..//.........AT.......`
Tanz 08-10		`......G.......C......//......C........//......T....` `..//.........AT.......`
Tanz 08-33		`......G.......C......//......C........//......T....` `..//.........AT.......`
Tanz 08-43		`......G.......C......//......C........//......T....` `..//.........AT.......`
Tanz 08-14	TZ-IV	`......G.......C......//......C........//......~....` `..//.........AT.......`
Tanz 08-02		`......G.......C......//......C........//......~....` `..//.........AT.......`
Tanz 08-01	TZ-V	`......G.......C......//......C........//......~....` `..//.........~~.......`
Tanz 08-44		`......G.......C......//......C........//......~....` `..//.........~~.......`
Tanz 08-11		`......G.......C......//......C........//......~....` `..//.........~~.......`
Tanz 08-05		`......G.......C......//......C........//......~....` `..//.........~~.......`

Figure 1. Symptoms of Lethal Decline, a Lethal Yellowing Type Syndrome of coconut in Tanzania. (Michel Dollet).
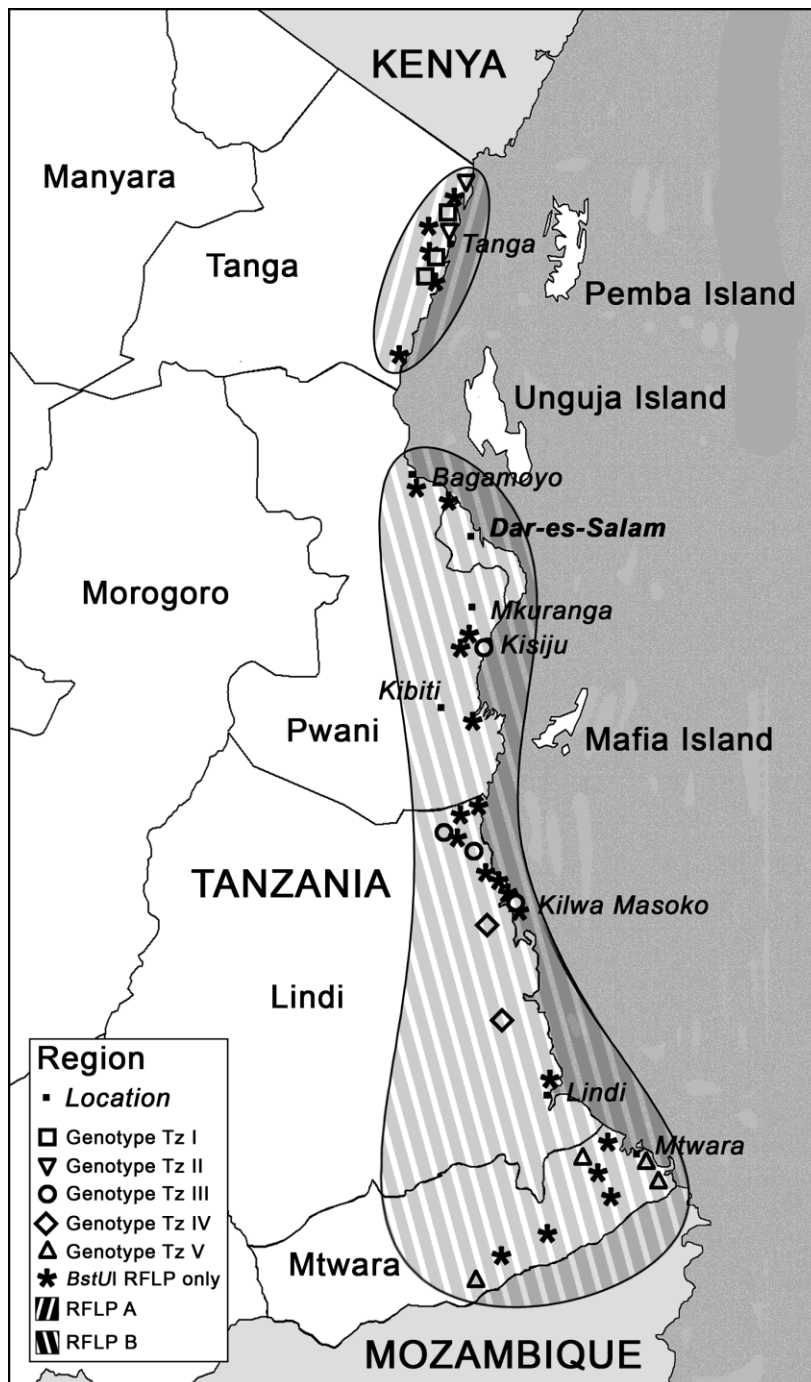
Figure 2 Geographical distribution of the LDT strains / clusters and phytoplasma genotypes
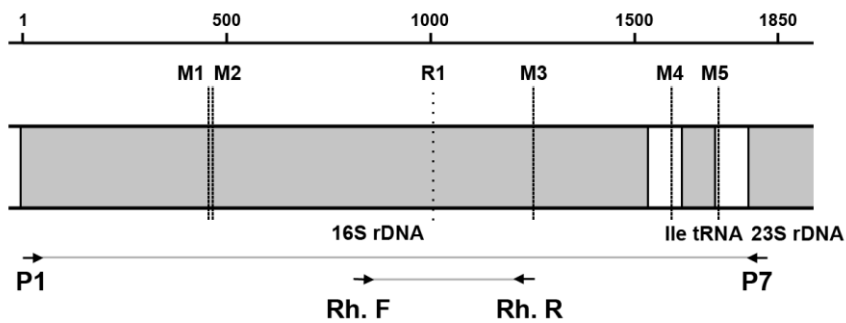
Figure 3 Diagrammatic representation of the 16S-23S rRNA gene organization in phytoplasma with the identified mutations and deletions in the various LDT isolates.
M1 and M2: double mutation characterizing the genotype TZ-I.
M3: mutation inside the Rhode's primers product (at the 5'extremity of the Reverse primer) defining genotype TZ-III.
M4: deletion inside the ITS1defining the TZ-IV genotype
M5: double deletion localized in the ITS2, separating out genotype TZ-V.
R1: mutation inside the Rhode's primers product for the isolate Tanz 08-06
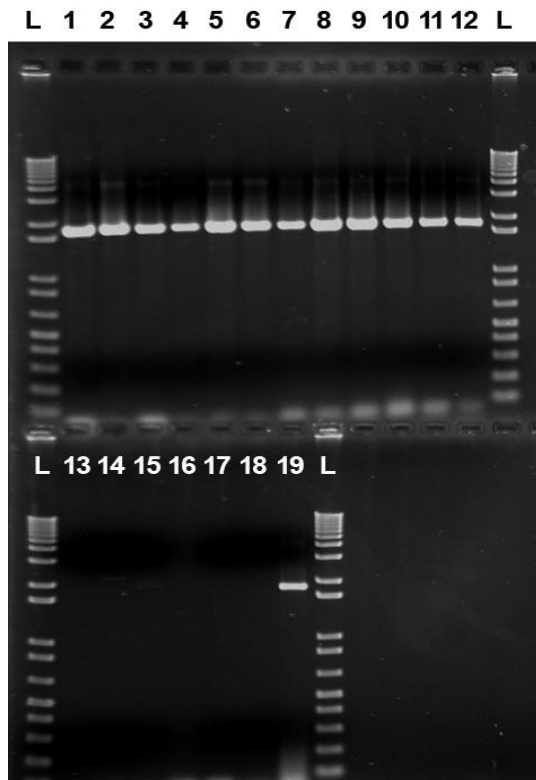
Figure 4 Amplification of rDNA from LDT-infected East African Tall coconuts using the phytoplasma universal primer pair P1 / P7. DNA for samples 1-12 were extracted from young inflorescence tissue, and for sample no. 13-17 from the spear leaves.

1 Tanz 08-01 (Mnawene, Mtwara), 2 Tanz 08-02 (Mirumba, Kilwa), 3 Tanz 08-04 (Huruma, Tandahimba), 4 Tanz 08-05 (Mpapura, Mtwara)), 5 Tanz 08-06 (Mtoni, Kilwa), 6 Tanz 08-09 (Mnalale, Newala), 7 Tanz 08-10 (Kikanda, Kilwa), 8 Tanz 08-11 (Madimba, Mtwara), 9 Tanz 08-14 (Kitomanga, Lindi), 10 Tanz 08-16 (Kisimatui, Tanga), 11 Tanz 08-18 (Mpirani Gezani, Mkinga) 12 Tanz 08-19 (Mnazi Mmoja, Pangani), 13 Tanz 08-45 (Mikinguni, Pangani), 14 Tanz 08-56 (Lupedile, Kilwa), 15 Tanz 08-76 (Sotele, Mkuranga), 16 Tanz 08-87 (Mwalusembe, Mkuranga), 17 Tanz 08-90 (Kigombe, Pangani), 18 Sterile distilled water 19 Clover Phyllody DNA, L 1 Kb Plus DNA Ladder
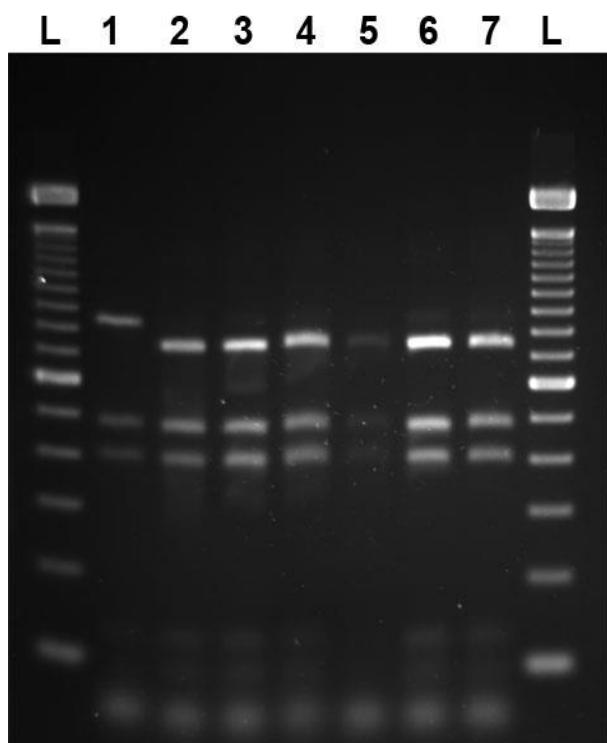
Figure 5 Agarose gel (2%) showing the two RFLP patterns obtained by digestion using *Bst*UI restriction endonuclease of P1/P7 PCR products, encompassing the 16S rRNA gene and the 16S-23S rRNA intron. Lane 1, RFLP A group, Tanz 08-19 (Mnazi Mmoja, Pangani), Lanes 2-7, RFLP B group, Tanz 08-32 (Mzizima,Kilwa), Tanz 08-33 (Miyumbu, Kilwa), Tanz 08-39 (Kilwa Masoko, Kilwa), Tanz 08-42 (Kale, Kilwa), Tanz 08-44 (Mkunya, Newala), Tanz 08-50 (Chambezi, Bagamoyo), L : 100 bp Invitrogen

Figure 6 Dendrograms, constructed by the Neighbor-Joining method, showing the phylogenetic relationships amongst the Lethal Yellowing Type Syndrome (LYTS) phytoplasmas based on sequences of the 16S rRNA gene. GenBank accession numbers for previously published sequences are shown in ( ) alongside the names of the phytoplasmas. Bootstrap values greater than 50% (expressed as percentages of 1000 replications) are shown, and branch lengths are proportional to the number of inferred character state transformations. Bar, substitutions per base.