Genetic variation, phylogenetic relationship and spatial distribution of ‘Candidatus Phytoplasma ulmi’ strains in Germany

B. Schneider1, B. Hüttel2, C. Zübert3 & M. Kube3

A recent survey in Germany revealed the wide presence of ‘Candidatus Phytoplasma ulmi’ in native elm stands. Accessions were studied for their genetic variability and phylogenetic relationship based on the conserved groEL and the variable imp gene. While the groEL sequences revealed a high intraspecific homology of more than 99%, the homology of the imp gene dropped to 71% between distantly related sequences. Twenty-nine groEL and 74 imp genotypes were distinguished based on polymorphic sites. Phylogenetic analysis of the groEL gene clustered all ‘Ca. P. ulmi’ strains and separated them from related phytoplasmas of the 16SrV group. The inferred phylogeny of the imp gene resulted in a different tree topology and separated the ‘Ca. P. ulmi’ genotypes into two clusters, one closely related to the flavescence dorée phytoplasma strain FD-D (16SrV-D), the other affiliated with the flavescence dorée phytoplasma strains FD-C and FD70 and the alder yellows phytoplasma (16SrV-C). In both phylograms, ‘Ca. P. ulmi’ genotypes from Scots elm trees formed a coherent cluster, while genotypes from European white elms and field elms grouped less strictly. The regional distribution pattern was congruent for some of the groEL and imp genotypes, but a strict linkage for all genotypes was not apparent.

Phytoplasmas are obligate vector-borne bacterial parasites associated with diseases of more than 1000 plant species1,2. They form the monophyletic taxon ‘Candidatus Phytoplasma’ and are members of the class Mollicutes3. ‘Candidatus Phytoplasma ulmi’ belongs to a diverse cluster of related phytoplasmas within this taxon commonly referred to as ‘elm yellows’, or the ‘16SrV group’. Members of this group are homogeneous in respect to their 16S rRNA sequence but can be divided into subgroups A to E based on restriction fragment length polymorphism analyses. In regard to affected plant species, insect vectors and ecological niches, the group is heterogeneous4. Beside the eponymous elm yellows, plant diseases such as flavescence dorée, rubus stunt, alder yellows and jujube witches’ broom are associated with phytoplasmas in this group5–4.

‘Ca. P. ulmi’ is associated with elm yellows, a serious disease of elms with a high mortal impact on populations, particularly in the eastern United States, where it was first reported in the last century7. Since then, elm yellows has been reported from a number of countries. Italy was the first country in Europe, but others followed8–14. In Germany, the disease was first reported from a Scots elm tree displaying witches’ broom symptoms11. This singular finding, however, gave no information on the general occurrence of the disease and the distribution of the pathogen in the German territory. It took more than 20 years until a local survey demonstrated the presence of ‘Ca. P. ulmi’ in European white elm trees in East Germany13, and the fact that all phytoplasma-positive trees lacked typical disease symptoms prompted a nationwide survey in native elm habitats and revealed an average infection rate of 28%15. This study also corroborated the observation of the previous survey that symptoms are rare, a circumstance explaining why the infections went unnoticed for such a long time. The situation in southern Europe is different. There, ‘Ca. P. ulmi’ infections seem to be associated generally with typical disease symptoms like witches’ brooms, leaf yellowing and premature leaf fall16,17, albeit the typical destructive phloem necrosis that American elm species display is not present in European elm species18,19. This decisive difference and the lack of molecular evidence were the determining factors in classifying this phytoplasma as a quarantine pathogen for the EPPO region20. Information on the distribution of elm yellows in Europe, which accumulated

1Thuenen-Institute of Forest Genetics, 15377 Waldsieversdorf, Germany. 2Max Planck Institute for Plant Breeding Research, Max Planck Genome Centre Cologne, 50829 Cologne, Germany. 3Department of Integrative Infection Biology Crops-Livestock, University of Hohenheim, 70599 Stuttgart, Germany. *email: bernd.schneider@thuenen.de
in the meantime, and the accruing molecular evidence of genetic relatedness between American and European ‘Ca. P. ulmi’ strains, prompted the EU to offset this classification for continental Europe21.

The first molecular evidence of a close relationship between American and European strains came from Southern blot hybridisation experiments using random DNA probes1, supported by restriction fragment length polymorphism (RFLP) and sequence analyses of the 16S rRNA gene thereafter4. Sequence homology of the 16S rRNA gene, together with particular features of the plant host and insect vector, led to the proposition of the novel taxon ‘Candidatus Phyttoplasma ulmi’1,4. However, a closer look at genetic homogeneity with less conserved genes such as map, uvrB-degV, secY and rpl22-rps3 revealed a greater intraspecific heterogeneity than expected from 16S rRNA data. A Serbian study of ‘Ca. P. ulmi’ accessions distinguished five genotypes with significant differences in relation to the American-type strain EY15 based on map, secY and rpl22-rps3 genes22. A combination of polymorphic sites of the same genes even distinguished 30 genotypes in a Croatian study17. Beside the taxonomically fundamental 16S rRNA gene, the groEL gene was proposed as an additional marker to resolve taxonomic issues in complex bacterial groups which has the ability to ensure accurate predictions and to replace more time intense multiple gene analyses. It has been demonstrated that a 550 bp groEL fragment can predict relationships within Thermoanaerobacter species with the same accuracy compared to a combinatorial analysis of three genes23. Sequence data for this gene were successfully used to differentiate subgroups further within the aster yellows, or 16Sr1 group, and the Bermuda grass white leaf, or 16SrXIV group22,25. A recently installed Web interface enables the comparison of groEL sequences from different taxonomic phytoplasma entities17, although the numbers of data are still low. In contrast to conserved genes, sequence-variable genes like the immunodominant membrane protein (imp) gene have only been used to a limited extent in the 16SrII (‘Ca. P. aurantifolia’)22 or 16SrX (apple proliferation)22–24 groups to resolve intraspecific variation, phylogenetic linkages and geographical distribution.

In this study, we examine genetic variations, the phylogenetic relationship and the regional distribution of ‘Ca. P. ulmi’ accessions with a groEL gene fragment and the imp gene—two markers of different resolving powers adding specific information to survey results published recently by the authors.

Results

DNA amplification of groEL- and imp fragments. The selected forward and reverse primers for the groEL- and imp genes amplified fragments of about 880 bp and 675 bp, respectively, from all 288 ‘Ca. P. ulmi’ accessions (Fig. 1). The primers also amplified groEL- and imp fragments from the phytoplasmas alder yellows strain ALY, the ‘Ca. P. ulmi’ strain ULW and the flavescence dorée strain FD70.

Genetic variation in the ‘Ca. P. ulmi’ groEL gene sequences. Homology among the ‘Ca. P. ulmi’ groEL sequences, including the ‘Ca. P. ulmi’ strain ULW, ranged between 99 and 100%, while homology to ALY (MT638097), FD70 (MT638098) and ‘Candidatus Phyttoplasma ziziphi’ sequences ranged between 94% and 96.4%. Accession 4120_UL_SN (MT638095), isolated from a European white elm tree, shared a lower sequence homology (95.7% to 96.4%) to all other ‘Ca. P. ulmi’ sequences and was closer related to the ALY (99.1%) and FD70 (98.6%) sequences. This accession was included in the phylogenetic analysis but not considered as a sensu stricto member of the species ‘Ca. P. ulmi’. Homology to the distantly related ‘Ca. Phyttoplasma mali’ groEL sequence was on average 69.7%.

The multiple alignment of the 790 bp groEL fragment from the ‘Ca. P. ulmi’ sequences, excluding sequence 4120_UL_SN, revealed a total of 18 polymorphic sites, of which 12 caused a nonsynonymous substitution (Table 1). However, mutations altering the amino acid in a significant number of isolates were only present at positions 18, 249 and 438. The remaining non-silent mutations occurred in genotype groups with a low number of isolates.

A comparison of the 288 groEL sequences, including accession 4120_UL_SN, on the basis of complete sequence homology, differentiated 29 genotypes, of which 18 occurred more than once (Table 2). The genotype

Figure 1. Amplification of imp (upper panel) and groEL (lower panel) fragments from ‘Ca. P. ulmi’ accessions. Only examples are shown. Sample names are indicated on the top, including the host plant and Federal State. Ug_h, DNA from a healthy Scots elm accession. The gel was photographed with a VWR Imager 2 System, Software Version 1.5.6.0 (https://de.vwr.com) and cropped with Photoshop CS3 (www.adobe.com).
groups correlated in most cases with the elm species from which they were isolated. In cases where the ‘Ca. P. ulmi’ genotype was associated with a different elm species, more than one elm species grew at the sampling location. However, if fruits were unavailable or leaf shapes difficult to interpret, a misidentification of the elm species cannot be excluded.

Table 1. Nucleotide position and variation thereof in the 790 bp groEL fragment of ‘Ca. P. ulmi’ accessions. 

| Position | Reference nucleotide and amino acid | A | C | G | T |
|----------|------------------------------------|---|---|---|---|
| 18       | A (N) – 1 (H) – – 11 (Y) |
| 26       | T – – 81 – – |
| 47       | A – – – 81 |
| 71       | C 105 – – 39 – |
| 102      | G (S) 1 (N) – – – |
| 224      | A – – 88 – |
| 249      | G (A) – – – 94 (S) |
| 262      | C (T) 2 (K) – – – |
| 290      | C – – – 67 |
| 342      | A (Y) – – 2 (C) – |
| 399      | G (D) 2 (N) – – – |
| 438      | T (S) 114 (T) – – – |
| 458      | C (F) – – 4 (I) – |
| 538      | C (A) – – 2 (V) – |
| 723      | T (M) – 4 (T) – – |
| 767      | C 1 – – 108 |
| 770      | A (Q) – – – 2 (H) |
| 771      | A (I) – – – 2 (F) |

Table 2. ‘Ca. P. ulmi’ genotypes distinguished based on complete groEL fragment homology. 

| Genotype/sum | Prevailing elm species | No. of sequences | Acc. No. | Other elm species |
|--------------|------------------------|------------------|----------|-------------------|
| I            | U. glabra              | 67               | MT638090 | U. laevis – 1     |
| II           | U. minor               | 61               | MT418907 | U. laevis 2       |
| III          | U. glabra              | 36               | MT638073 | U. minor 1        |
| IV           | U. laevis              | 35               | MT638072 | U. glabra 1       |
| V            | U. laevis              | 34               | MT638096 | U. minor 3       |
| VI           | U. glabra              | 6                | MT638095 | U. glabra – –     |
| VII          | U. glabra              | 6                | MT638071 | U. laevis – 1    |
| VIII         | U. glabra              | 5                | MT630879 | U. glabra – –     |
| IX           | U. laevis              | 4                | MT638088 | U. laevis – 1    |
| X            | U. minor               | 4                | MT638091 | U. minor 2        |
| XI           | U. laevis              | 4                | MT638092 | U. laevis – –     |
| XII          | U. glabra              | 3                | MT638078 | U. glabra – –     |
| XIII         | U. glabra              | 3                | MT638069 | U. glabra – –     |
| XIV          | U. glabra              | 2                | MT638094 | U. glabra – –     |
| XV           | U. glabra              | 2                | MT638074 | U. glabra – –     |
| XVI          | U. minor               | 2                | MT638083 | U. minor – –      |
| XVII         | U. glabra              | 2                | MT638087 | U. glabra – –     |
| XVIII        | U. glabra              | 2                | MT638076 | U. glabra – –     |
| Sum          | 278                    | 8                | 8        | 13                |

Table 2. ‘Ca. P. ulmi’ genotypes distinguished based on complete groEL fragment homology. 

| Genotype/sum | Prevailing elm species | No. of sequences | Acc. No. | Other elm species |
|--------------|------------------------|------------------|----------|-------------------|
| I            | U. glabra              | 67               | MT638090 | U. laevis – 1     |
| II           | U. minor               | 61               | MT418907 | U. laevis 2       |
| III          | U. glabra              | 36               | MT638073 | U. minor 1        |
| IV           | U. laevis              | 35               | MT638072 | U. glabra 1       |
| V            | U. laevis              | 34               | MT638096 | U. minor 3       |
| VI           | U. glabra              | 6                | MT638095 | U. glabra – –     |
| VII          | U. glabra              | 6                | MT638071 | U. laevis – 1    |
| VIII         | U. glabra              | 5                | MT630879 | U. glabra – –     |
| IX           | U. laevis              | 4                | MT638088 | U. laevis – 1    |
| X            | U. minor               | 4                | MT638091 | U. minor 2        |
| XI           | U. laevis              | 4                | MT638092 | U. laevis – –     |
| XII          | U. glabra              | 3                | MT638078 | U. glabra – –     |
| XIII         | U. glabra              | 3                | MT638069 | U. glabra – –     |
| XIV          | U. glabra              | 2                | MT638094 | U. glabra – –     |
| XV           | U. glabra              | 2                | MT638074 | U. glabra – –     |
| XVI          | U. minor               | 2                | MT638083 | U. minor – –      |
| XVII         | U. glabra              | 2                | MT638087 | U. glabra – –     |
| XVIII        | U. glabra              | 2                | MT638076 | U. glabra – –     |
| Sum          | 278                    | 8                | 8        | 13                |
Genetic variation in the 'Ca. P. ulmi' imp gene sequences. The length of the 'Ca. P. ulmi' imp gene varied. While the majority of isolates (N = 253) showed a gene length of 465 bases, 34 isolates had a gene six bases larger. The imp gene of accession 4120_Ul_SN (MT668488) was shorter by six bases, a feature shared with the phytoplasma strain ALY. Homology among the 'Ca. P. ulmi' imp gene sequences ranged from 71 to 100%, whilst homology to the imp gene of the 16SrV-D group phytoplasma FD-D ranged between 66 and 79%, and homology to the 16SrV-C group sequences of ALY (MT668499), FD70 (MT668500) and 'Ca. P. ziziphi' was generally lower with a narrower bandwidth and ranged between 61 and 65%. The accession 4120_Ul_SN shared, as evidenced with the groEL gene, higher homology with 16SrV-C phytoplasmas (ALY, 83% and FD70, 92%) than with the 'Ca. P. ulmi' accessions (≈ 71%). Homology of the 'Ca. P. ulmi' strains to the distantly related species 'Ca. P. mali' was about 42%.

The majority of 'Ca. P. ulmi' accessions (N = 254), the strain ULW and FD70 coded for a protein of 154 amino acids (aa). Thirty-four accessions, mostly from European white elms (N = 26), including the strain FD-D, encoded a protein two aa larger. The imp gene of accession 4120_Ul_SN coded for a protein of 152 aa only, a feature shared with ALY. Due to the high sequence heterogeneity between the 'Ca. P. ulmi' protein sequences, homology between distantly related isolates dropped to 48%. A protein-based reassessment of the genotype groups revealed that three sequences, which showed a single base mutation, shared identical IMP sequences to others reducing the number of genotypes to 71. Homology to the 16SrV-C group phytoplasmas ALY and FD70 Figure 2. Alignment of imp sequences from 4319_Ug_SN (MT668492), 3486_Ug_ST (MT668480), 0371_UL_BB (MT668429), ULW (MT418908) and the 16SrV-group phytoplasmas ALY (MT668499), FD70 (MT668500) and 'Ca. P. ziziphi' (CP025121) with Clustal W. The consensus sequence is given above the alignment, showing bases represented in more than 50% of sequences. Sequence alignments have been performed with Clustal W (www.ebi.ac.uk/clustalw).
ranged from 42 to 52%, to the 16SrV-D group phytoplasma FD-D from 45 to 61% and decreased below 39% in comparison to the 16SrV-B group phytoplasma 'Ca. P. ziziphi'. The accession 4120_Ul_SN shared homology of 34%, 54% and 86% to the 16SrV-B, -D and C phytoplasmas, respectively.

'Ca. P. ulmi' phylogeny based on groEL sequences. The phylogeny inferred on the groEL gene indicated a monophyletic origin of the 'Ca. P. ulmi' accessions, except for genotype 4120_Ul_SN, which clustered with ALY and FD70 (Fig. 3). The 'Ca. P. ulmi' clade split into two sub-branches whereby the genotypes from Scots elms and European white elms formed a separate entity, and the genotypes from field elm branched off from a Scots elm sub-clade. In a preliminary analysis, two groEL sequences (KJ939991, KJ939992) assigned to the species 'Ca. P. ulmi' clustered with the ALY- and FD70 groEL sequences but not with the 'Ca. P. ulmi' sequences obtained in this work. The two sequences were not included in the final analysis, as the fragment size of ca. 550 nt reduced the number of discriminating positions and affected the tree topology (data not shown).

'Ca. P. ulmi' phylogeny based on imp sequences. The phylogeny inferred on the imp gene sequences divided the 'Ca. P. ulmi' genotypes into two clades. Genotypes of each clade were affiliated with phytoplasmas of a different 16SrV subgroup (Fig. 4). The majority (N = 238) of 'Ca. P. ulmi' sequences clustered with the flavescence dorée strain FD-D, a member of the 16SrV-D subgroup. Genotypes of this clade split further into three branches comprising, with few exceptions (Fig. 4, red arrows), isolates of one elm species (Fig. 4, upper part). Fifty 'Ca. P. ulmi' genotypes were more closely related to phytoplasmas of the 16SrV-C subgroup, represented by the flavescence dorée strains FD70 and FD-C and the alder yellows phytoplasma ALY (Fig. 4, bottom part). These 'Ca. P. ulmi' genotypes separated further according to their gene length and were less strictly arranged according to the elm species of origin. The phylogenetic tree inferred from deduced protein sequences resembled the DNA-derived tree, except that the 16SrV-C subgroup phytoplasmas neighboured the 16SrV-D group strain (data not shown).

Table 3. 'Ca. P. ulmi' groups established on complete homology of the imp gene. aOnly genotypes which occurred at least twice are listed. bAccession number of a representative group member. cNumber of isolates not associated with the prevailing elm species. dThe number includes 'Ca. P. ulmi' strain ULW.

| Genotype/sum | Prevailing elm species | No. of sequences | Acc. No. a | Other elm species b |
|--------------|------------------------|-----------------|------------|---------------------|
| I            | U. glabra              | 101             | MT668492   | U. glabra U. laevis U. minor |
| II           | U. minor               | 26              | MT418908   | 1 3 –                |
| III          | U. glabra              | 15              | MT668459   | – – –                |
| IV           | U. laevis              | 14              | MT668429   | 2 – 1                |
| V            | U. laevis              | 9               | MT668462   | – – 3                |
| VI           | U. laevis              | 9               | MT668464   | 1 – –                |
| VII          | U. laevis              | 9               | MT668445   | – – 1                |
| VIII         | U. minor               | 9               | MT668456   | – 1 –                |
| IX           | U. minor               | 8               | MT668475   | – – –                |
| X            | U. glabra              | 4               | MT668444   | – – –                |
| XI           | U. glabra              | 4               | MT668434   | – – 1                |
| XII          | U. minor               | 3               | MT668437   | – – –                |
| XIII         | U. laevis              | 2               | MT668453   | – – 1                |
| XIV          | U. laevis              | 2               | MT668483   | – – –                |
| XV           | U. glabra              | 2               | MT668435   | – – –                |
| XVI          | U. glabra              | 2               | MT668497   | – – –                |
| XVII         | U. glabra              | 2               | MT668482   | – – –                |
| XVIII        | U. glabra              | 2               | MT668494   | – – –                |
| XIX          | U. glabra              | 2               | MT668458   | – – –                |
| XX           | U. minor               | 2               | MT668446   | – – –                |
| XXI          | U. laevis              | 2               | MT668498   | – – –                |
| XXII         | U. glabra              | 2               | MT668447   | – – –                |
| XXIII        | U. minor               | 2               | MT668485   | – – –                |
| XXIV         | U. laevis              | 2               | MT668440   | – – –                |
| XXV          | U. minor               | 2               | MT668442   | – – –                |
| XXVI         | U. minor               | 2               | MT668435   | – – 1                |
| XXVII        | U. glabra              | 2               | MT668448   | – – –                |
| XXVIII       | U. glabra              | 2               | MT668480   | – – –                |
| Sum          |                        | 242             | 4 13 11    |                     |

www.nature.com/scientificreports/
Non-synonymous vs. synonymous mutations. The ratio between non-synonymous over synonymous mutations was 1.53 (dN = 0.23; dS = 0.15) for imp and 1.00 (dN = 0.01; dS = 0.01) for the groEL-fragment. The ratio of dN/dS > 1 for the imp gene indicates a positive selection, which was not the case for the groEL-fragment.

Spatial distribution pattern of groEL- and imp genotypes and gene linkage. The spatial distribution of groEL and imp genotypes was not homogeneous in the territory, and flexibility in the combination of genotype linkages was apparent. The polymorphism of both genes resulted in genotype groups with a low number of datasets, which were insufficient for conclusive results. Therefore, only the first five genotype groups comprising nine sequences or more were analysed (see Tables 2 and 3). The spatial distribution of the groEL- and imp genotypes I to V is displayed in Fig. 5, and while the groEL genotypes I and III were present in South, Central and East Germany, genotypes II and IV were absent in the central and the south-eastern regions (Fig. 5, top panel). GroEL genotype V sequences were only found in East Germany.

The distribution of the imp genotypes I, II and V largely matched the respective groEL genotype spread (Fig. 5, bottom panel), although small differences existed. The imp genotype I for example, was in contrast to the groEL genotype I, also present along the upper Rhine valley, in the far south-west of Germany. A greater disparity of distribution was observed between the ‘Ca. P. ulmi’ imp and groEL genotype groups III and IV. The imp genotype III was absent in East Germany, while the imp genotype IV did not occur in the south of the country.

The regional distribution of genotypes was largely congruent with the distribution of the elm species. Genotype groups I and III were mostly associated with Scots elms, genotype II with field elms and genotypes IV and V with European white elms. However, there was some flexibility in the linkage of groEL- and imp genotypes. The imp genotype I, comprising 101 sequences, was linked to 15 different groEL genotype groups, of which the groEL...
Figure 4. Evolutionary analysis by maximum likelihood method of the *imp* gene from 74 *Ca. P. ulmi* genotypes and the 16SrV group phytoplasmas flavescence dorée strains FD-C, FD-D and FD70, ALY and *Ca. P. ziziphi*. All sequences, except those from *Ca. P. ulmi* strain ULW, FD-C, FD-D and *Ca. P. ziziphi*, were obtained in this study. Isolate designation and additional information as given in Fig. 3. There were a total of 429 positions in the final dataset, and bootstrap values (> 50) for 1,000 replicates are indicated. Red arrows indicate accessions not originating from field elms. The phylogenetic tree was compiled with the software MEGA version X (www.megasoftware.net).
genotype group I comprised the most with 57 sequences, followed by group III with 16 sequences and group VI with six sequences. The remaining 22 sequences were part of 12 other \textit{groEL} genotype groups. The same was true when considering the \textit{groEL} group II in comparison with \textit{imp} genotype group II. All 26 sequences of the \textit{imp} genotype group II were part of the corresponding \textit{groEL} group, but the remaining 35 \textit{groEL} sequences were associated with a further 19 \textit{imp} genotype groups. Neither the amplification products nor the electropherograms of the \textit{groEL} or \textit{imp} sequences provided evidence that the amplimers consisted of polymorphic fragments. The observed gene combinations were therefore considered as a true linkage and the result of selective evolutionary pressure.

**Discussion**

A recently conducted survey demonstrated the presence of \textit{’Ca. P. ulmi’} in many parts of Germany, but the genetic variation of the regional strains and their phylogenetic relationship were not examined. In this work, we address both issues through the sequence analyses of two genes. The conserved \textit{groEL} gene was selected to provide a better intraspecific resolution, due to a higher sequence variation compared to the 16S rRNA gene\textsuperscript{24,25,30}, and the variable \textit{imp} gene was chosen to tap the full range of genetic polymorphism of a variable gene.

The primer sets for amplifying both genes were deduced from the database entries of the respective genes from the \textit{’Ca. P. ulmi’} strain ULW. The oligonucleotides were compared to target genes of close relatives from the 16SrV-B, -C and -D subgroups, where available. Due to the high conservation of the \textit{groEL} gene and the likelihood of an unspecific amplification of other bacterial \textit{groEL} genes from crude plant extracts, the reverse primer showed six mismatches to the \textit{’Ca. P. ziziphi’} sequence, the only other full-length \textit{groEL} gene in the 16SrV group published. Primers for the amplification of the \textit{imp}-\textit{pyrG} region were fully homologous to the respective sequences of \textit{’Ca. P. ulmi’} strain ULW and the \textit{flavescence dorée} strain FD-C. Both primer sets enabled the amplification of the target fragments from all \textit{’Ca. P. ulmi’} accessions, including the strains FD70 and ALY. Although, the sequencing results revealed no evidence of mixed infection of plants by more than one \textit{’Ca. P. ulmi’} strain, we do not exclude this possibility. It is conceivable that other \textit{’Ca. P. ulmi’} strains were present, but they must have been present in small numbers so alternative base calls were below the threshold line of the electropherogram. A closer examination of this matter would have exceeded the scope of this work.

Previous studies have demonstrated the genetic variability of \textit{’Ca. P. ulmi’} isolates, but genotype classification was based on a combined evaluation of three to four genes\textsuperscript{17,22}. In this work, we examined for the first time the resolving power of the \textit{groEL} gene for intraspecific \textit{’Ca. P. ulmi’} differentiation. Although sequence polymorphism did not exceed 1%, except for accession 4120_Ul_SN, 29 genotypes were identified. Compared to the intraspecific 16S rDNA sequence heterogeneity of < 0.4\%, the sole consideration of the \textit{groEL} gene allowed a significantly higher intraspecific resolution. It is conceivable that an even greater number of genotypes would have been

![Figure 5. Spatial distribution patterns of \textit{groEL} genotypes I to V (upper panel) and \textit{imp} genotypes I to V (bottom panel). The number of sequences per genotype is indicated and illustrated as coloured dots. Due to the map scale, not all dots are visible. Maps were created with the software Garmin BaseCamp version 4.7.1 (www.garmin.com). Individual maps were assembled and labelled with the software program Photoshop CS3 (www.adobe.com).](image-url)
identified, as the examined accessions represented only 16% of the ‘Ca. P. ulmi’-positive samples identified in the previous survey. However, samples were carefully selected based on an even geographical spread and on a representative number of samples from all three elm species.

The groEL and imp genotype groups comprised, with a few exceptions, accessions originating from the same elm species, although parallel genotype groups with the same host species existed. Taking into consideration the many sequences of the larger genotype groups, the distinct host–pathogen strain association seems not to be arbitrary. It is unlikely that the groEL gene product, a chaperonin, is responsible for this specificity; however, it is likely that groEL is correlated to one or more such genes specifying this feature.

One of these candidate genes could be imp. The gene, but in particular the encoded protein, is a focus of phytoplasma research and the subject of host–pathogen interaction studies and the development of immunodiagnostic tools. However, despite its variability, the gene has been rarely used for intraspecific sequence polymorphism studies. In this study, we have employed for the first time the imp gene of ‘Ca. P. ulmi’ for this purpose and determined a high strain polymorphism, which was not only expressed by its sequence length, but also by its sequence composition. A calculated overall dN/dS ratio of 1.53 indicates that the ‘Ca. P. ulmi’ imp gene is subject to a high selective pressure, a factor that was demonstrated to act on other phytoplasma membrane protein genes. The high number of imp genotypes identified herein point in this direction, but on the other hand, the majority of sequences belonged to a few groups, and again, those groups were, as for the groEL gene, associated with a specific elm species. Whether or not the ‘Ca. P. ulmi’ membrane protein has an important function in host–pathogen recognition, as studies with other phytoplasmas suggest, remains to be demonstrated, but in a recent study the interaction of the FD-D IMP with gut membrane proteins of vector aceF-based on the imp obtained after comparing the misidentification. The breakup of a monophyletic relationship is not unique for ‘Ca. P. ulmi’ accessions, the loss of discriminating characters changed the tree hierarchy. This example demonstrates that although the groEL gene is a proven universal target for phylogenetic analysis, a certain sequence length might be necessary to exploit the full information content of a gene. If not so, biologically significant features might be obstructed, resulting in a less meaningful phylogenetic tree.

The phylogenetic analysis of the groEL fragment placed all, except one ‘Ca. P. ulmi’ accession, on one root separated from members of the 16SrV-B and C subgroup. The accessions on this root were arranged according to the elm species from which they were isolated. This highly ordered elm species-related grouping broke up when shorter sequences (<560 bp) were included in the analysis, and although these sequences did not cluster with the ‘Ca. P. ulmi’ accessions, the loss of discriminating characters changed the tree hierarchy. This example demonstrates that although the groEL gene is a proven universal target for phylogenetic analysis, a certain sequence length might be necessary to exploit the full information content of a gene. If not so, biologically significant features might be obstructed, resulting in a less meaningful phylogenetic tree.

The imp-derived phylogenetic tree was more complex and separated the ‘Ca. P. ulmi’ accessions on two separate roots. Interestingly, the accessions on both roots behaved differently. Whereas the 16SrV-D-related strains clustered according to their elm host, the 16SrV-C-related accessions were grouped according to their sequence length, rather than to their elm host. Although, the Scots elm accession and the European white elm accession which were interspersed in the 16SrV-C-related field elm (Ulmus minor) branch might have been a matter of species misidentification. The breakup of a monophyletic relationship is not unique for ‘Ca. P. ulmi’. A similar result was obtained after comparing the imp genes of hypovirulent ‘Ca. P. prunorum’ strains, which displayed monophyly based on the aceF-gene. This is, however, not always the case as shown for the stolbur membrane protein gene stamp. Here, all stamp sequences grouped according to established tuf clusters. Why the immunodominant membrane protein genes of some phytoplasma groups seem to evolve independently from other genes might be explained by their interaction with host proteins, but this does not seem to apply for the imp sequences of the flavescence dorée strains FD-C and FD-D, which differ but share the same plant and insect hosts.

We tried to identify a superordinate clustering principle, like provenience, for arranging genotypes at the terminals of a clade. For some, like the European white elm accessions (N = 29) in the upper part of the phylogram (Fig. 4), their origin laid in the east of Germany, but for other clades and sister clades, accessions of all states assembled. Interesting to note was the position of the three flavescence dorée strains, FD70, FD-C and FD-D, relative to the ‘Ca. P. ulmi’ accessions. FD strains have been distinguished on the basis of the map and uvr-degV genes and their geographic origin. Here, we could show that the imp gene of FD70, a strain from south-western France, was identical to the imp gene of FD-C, a strain restricted to Italy. Both strains were closely related to alder yew, to accession 4120_Ul_SN and to a subset of ‘Ca. P. ulmi’ accessions mainly isolated from European white elm. The imp gene of the FD-D strain, on the other hand, was more closely related to ‘Ca. P. ulmi’ strains than to FD70 and FD-C. This situation demonstrates the range of genetic diversity within the 16SrV group and the complex phylogenetic relationship between its members. However, the fact that a variable gene, like imp, is more related between otherwise more remote strains highlights its biological importance.

The study revealed a high number of groEL- and imp genotype groups, many of which comprised too few sequences to expect significant results. The major genotype groups, however, showed distinct distribution patterns, and in many cases these patterns corresponded, albeit a certain number of accessions, either in the groEL- or imp genotype groups, were classified in non-corresponding groups. Hence, a certain flexibility of gene linkage exists, but not to a degree of complete randomness. This is evidenced by the fact that such groEL- or imp sequence variants were always associated with other imp- or groEL sequence variants of the same elm species. It also seems that another so far unknown factor determines the host–pathogen association. The delineation of elm species-specific phytoplasmas in regions with mixed elm populations points to an insect vector as a decisive factor;
Indeed, monophagous phloem feeders have been identified. For example, for field elm, *Macropsis glandacea*, (syn. *Macropsis mendax*) the only experimentally confirmed vector for ‘Ca. *P. ulmi*’ in Italy, and *lassus scutellaris* have been described[41]. Furthermore, a number of polyphagous phloem-feeders such as *Allygidiis atomaria* or *Empoasca vitis* have been described, but for many *Auchenorrhyncha* species found on elm trees, the host specificity is not completely clear. Therefore, the initiative to identify ‘Ca. *P. ulmi*’ vectors might provide important new information on how the spatial distribution of ‘Ca. *P. ulmi*’ strains has evolved.

**Materials and methods**

**Plant samples, DNA extraction and phytoplasma strains.** Shoot samples from Scots elms (*Ulmus glabra*), European white elms (*Ulmus laevis*) and field elms (*Ulmus minor*) were collected from October 2017 to May 2019 in natural elm habitats in Germany. Total nucleic acids were extracted from 125 mg of phloem tissue, according to a procedure described by Ahrens and Seemüller[42], and 288 phytoplasma-positive samples from 133 sampling sites were selected for analysis. From 18, 94 and 21 sites, one, two and up to four accessions were examined, respectively. Identification of ‘Ca. *P. ulmi*’ was performed by a pathogen-specific quantitative real-time PCR assay described by Schneider et al.[15]. DNA from the 16SrV-C group phytoplasma alder yellows, strain ALY[46] and the 16SrV-A group phytoplasma ‘Ca. *P. ulmi*’ strain ULW (obtained by G. Morvan, INRA, France), maintained in *Catharanthus roseus*, was extracted from leaf midribs. DNA from the 16SrV-C group phytoplasma *flavescence dorée*, strain FD70[47] maintained in *Vicia faba*, was obtained from M. Maixner, JFK, Germany. The nucleic acid pellets were resuspended in 200 µl of ddH2O.

**Primer design.** The primer design for amplifying the ‘Ca. *P. ulmi*’ groEL- and imp fragments was based on the full-length sequence of the groEL (MT418907), dnaD (MT418910), imp (MT418908) and pyrG (MT418909) gene from the ‘Ca. *P. ulmi*’ strain ULW. Validation of the deduced primers was performed via alignment with the groEL gene of ‘Ca. *P. ziziphi*’ (CP025121) and the dnaD-imp-pyrG fragments of ‘Ca. *P. ziziphi*’ (CP025121) and of the *flavescence dorée* phytoplasma strain FD-C (KJ402359). All primer sequences are listed in Table 4.

**Polymerase chain reaction (PCR) amplification and sequencing.** PCR fragment amplification was carried out in 20 µl reactions, using 20 pmol of forward and reverse primer, 2 µl of DNA extract, 10 µl of 2× primaQuant PCR master mix (Steinbrenner, Wiesenbach) and 4 µl of sterile water. The amplification was performed in a qTower (Analytik Jena AG, Jena) with the following cycling conditions: three minutes at 95 °C for the initial denaturation and polymerase activation, followed by 35 cycles consisting of 10 s at 95 °C, 20 s at 59 °C ([groEL], or 54 °C ([imp]) and 45 s at 72 °C. The final elongation step was set to 5 min at 72 °C. After amplification, 2 µl of the PCR reaction mix were used to verify product yield and size by electrophoresis in a GelRed-stained 2% agarose gel. The PCR fragments of the remaining 18 µl were purified by a PCR purification kit (Qiagen, Hilden) and quantified using a Qubit fluorometer (Thermo Fisher Scientific, Darmstadt). Sequencing was performed with the forward and reverse primers used for amplification.

**Sequence analysis, phylogenetic tree construction and calculation of non-synonymous vs. synonymous mutations.** The forward and reverse sequences of the PCR fragments were aligned for sequence correspondence, using the program Seqman (DNASTAR INC, Madison). Consensus sequences were used for analysis when fully matching. In the case of ambiguous base calls, sequencing was repeated. The groEL fragments were trimmed to a length of 790 bp for a common 5′ start and 3′ end. The 675 bp imp/pyrG fragments, hereafter referred to as imp fragments or genes, were trimmed to the coding region of the imp gene. The groEL- and imp genes from the phytoplasma strains ALY, FD70 and ULW were also sequenced, trimmed as above and included in the analysis. To avoid multiple database entries with identical sequences, the DNA sequence of one representative of each genotype group was submitted to the Genbank database. The accession numbers for the groEL- and imp sequences are MT638069 to MT638098 and MT668426 to MT668500, respectively. The groEL- and imp gene sequences of ‘Ca. *P. ziziphi*’ (CP025121) and ‘Ca. *P. mali*’ (CU469464) were extracted from the full genome sequences. Multiple alignments were performed using Clustal W[44]. The phylogenetic analysis of the nucleotide sequences and amino acids sequences was performed with MEGA X software[45], using the maximum likelihood method and the complete deletion of positions containing gaps and missing data, and with a bootstrap value for 1000 replicates[46]. The phylogenetic analysis comprised 292 sequences for the groEL and 294 sequences for the imp genes. The database entries KJ939991 and KJ939992 of a partial groEL gene from 16SrV-group phytoplasmas, classified as ‘Ca. *P. ulmi*’, were included in a preliminary phylogenetic analysis. For imp tree construction, the database entries MK614707 (FD-D) and KJ402359 (FD-C Piemonte) were included. For clarity, the depicted phylogenetic trees were inferred with only one representative of each genotype group.

| Primer abbreviation | Sequence (5’–3’) |
|---------------------|-----------------|
| fEY_groEL           | GTTAATGATGGCGTTACAATCGC |
| rEY_groEL           | GTTAAAGAAGGACTTTATCCGC |
| fEY_imp             | CATTATTATATCTGTATATAAAC |
| rpyrG               | GACCTTTAAACGATCC |

Table 4. Amplification primers for the groEL- and the imp gene fragments.
The multiple sequence alignments of the groEL- and imp fragments were also used to calculate the ratio between non-synonymous and synonymous sites (dN/dS) which were calculated as an average over all codon sites in each gene using the Nei-Gojobori method with Jukes-Cantor model included in the MEGA X software package.

**Distribution pattern of groEL and imp genotypes and gene linkage.** Coordinates of individual trees were recorded in WGS84 format and imported to the software program BaseCamp version 4.7.1 (Garmin Ltd, Olathe). For a graphical display of spatial distribution, groEL- and imp genotypes were depicted as differently coloured dots on maps. groEL- and imp sequences obtained from the same DNA extract were considered as linked when sequencing electropherograms gave no indication of polymorphic PCR products. Furthermore, it was assumed that primer sets amplified the genes of the prevalent ‘Ca. P. ulmi’ strain and that both genes were present in single copy.

**Code availability**
Accession codes: DNA sequences determined in the course of this study have been deposited in GenBank (ncbi.nlm.nih.gov) under accession numbers MT638069 to MT638098 for the groEL fragments and MT668426 to MT668500 for the imp sequences.

Received: 1 July 2020; Accepted: 23 November 2020
Published online: 14 December 2020

**References**
1. Seemüller, E., Garnier, M. & Schneider, B. Mycoplasmas of plants and insects. In Molecular Biology and Pathology of Mycoplasmas (eds Razin, S. & Hermann, R.) 91–116 (Kluwer Academic/Plenum Publishers, London, 2002).
2. Sugio, A. M., Kingdom, H. N., Grieve, V. M., Manimekalai, R. & Hogenhout, S. A. Diverse targets of phytoplasma effectors: from plant development to defense against insects. *Ann. Rev. Phytopathol.* **49**, 175–195 (2011).
3. IRPCM Phytoplasma/Spiroplasma Working Team-Phytoplasma Taxonomy Group. ‘Candidatus Phytoplasma’, a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *Int. J. Syst. Evol. Microbiol.* **54**(4), 1243–1255 (2004).
4. Lee, I.-M., Martini, M., Marcone, C. & Zhu, S. F. Classification of phytoplasma strains in the elm yellows group (16SrV) and proposal of ‘Candidatus Phytoplasma ulmi’ for the phytoplasma associated with elm yellows. *Int. J. Syst. Evol. Microbiol.* **54**(2), 337–347 (2004).
5. Angelini, E., Clair, D., Borgo, M., Bertaccini, A. & Boudon-Padieu, E. Mycoplasmas of plants and insects. In *Molecular Biology and Pathology of Mycoplasmas* (eds Razin, S. & Hermann, R.) 91–116 (Kluwer Academic/Plenum Publishers, London, 2002).
6. Malembic-Maher, S. et al. Genetic diversity of European phytoplasmas of the 16SrV taxonomic group and proposal of ‘Candidatus Phytoplasma rubi’. *Int. J. Syst. Evol. Microbiol.* **61**(9), 2129–2134 (2011).
7. Mäurer, R., Seemüller, E. & Sinclair, W. A. Genetic relatedness of mycoplasma-like organisms affecting elm, alder, and ash in Europe and North America. *Phytopathology* **83**, 971–976 (1993).
8. Jung, H. Y. et al. ‘Candidatus Phytoplasma ziziphi’, a novel phytoplasma taxon associated with jujube witches’-broom disease. *Int. J. Syst. Evol. Microbiol.* **53**(4), 1037–1041 (2003).
9. Swingle, R. U. A phloem necrosis of elm. *Phytopathology* **28**(10), 757–759 (1938).
10. Goidànic, G. Glischòppzi dell’olmo. *Inf. Fitopatol.* **14**, 8 (1951).
11. Seemüller, E. Laubgehölzmycoplasmosen in Europa. *Nachrichtenblatt Pflanzenschutzdienst* **44**, 145–148 (1992).
12. Plese, N. & Juretic, N. Virus disease of field elm (*Ulmus minor Mill.*) in Croatia. *Šumarski List* **123**(3/4), 95–100 (1999).
13. Eisold, A.-M., Kube, M., Hols, S. & Bütten, C. First Report of ‘Candidatus Phytoplasma ulmi’ in *Ulmus laevis* in Germany. *Common. Agric. Appl. Biol. Sci.* **80**(3), 575–578 (2015).
14. De Jonghe, K., Deeren, A. M. G., Goedefroit, T. & Ronse, A. First report of ‘Candidatus Phytoplasma ulmi’ on elm in Belgium. *Plant Dis.* **103**(7), 1763 (2019).
15. Schneider, B., Kätzel, R. & Kube, M. Widespread occurrence of ‘Candidatus Phytoplasma ulmi’ in elm species in Germany. *BMC Microbiol.* **20**(1), 1–12 (2020).
16. Marcone, C., Raguzzino, A. & Seemüller, E. Identification and characterization of the phytoplasma associated with elm yellows in southern Italy and its relatedness to other phytoplasmas of the elm yellows group. *Eur. J. Forest. Pathol.* **27**(1), 45–54 (1997).
17. Katanic, Z., Krstic, L., Jezic, M., Zebec, M. & Curkovic-Perica, M. Molecular characterization of mycoplasmas infecting *Ulmus* species in Croatia and their impact on *Ulmus* spp. *Plant Pathol.* **65**(9), 1365–3059 (2016).
18. Braun, E. J. & Sinclair, W. A. Phloem necrosis of elms: symptoms and histopathological observations in tolerant hosts. *Phytopathology* **69**(4), 354–358 (1979).
19. Mitterperger, L. Elm yellows in Europe. In *The Elms: Breeding, Conservation, and Disease Management* (ed. Dunn, C. P.) 103–119 (Kluwer Academic Publisher, New York, 2000).
20. OEP/EPPPO. Data sheets on quarantine organisms No. 27, Elm phloem necrosis (mycoplasm). OEP/EPPPO Bulletin **9**(2), (1979).
21. Anonymous. Regulation (EU) 2016/2031 of the European Parliament of the Council of 26 October 2016 on protective measures against pests of plants, amending Regulations (EU) No 228/2013, (EU) No 652/2014 and (EU) No 1143/2014 of the European Parliament and of the Council and repealing Council Directives 69/464/EEC, 74/674/EEC, 93/85/EEC, 98/57/EC, 2000/29/EC, 2006/91/EC and 2007/33/EC. OF J L **317**, 4–104 (2016).
22. Jović, J. et al. Multigene sequence data and genetic diversity among ‘Candidatus Phytoplasma ulmi’ strains infecting *Ulmus* spp., *Serbia. Plant Pathol.* **60**(2), 356–368 (2011).
23. Verbeke, T. J. et al. Predicting relatedness of bacterial genomes using the chaperonin-60 universal target (cpn60 UT): application to *Thermoanaerobacter* species. *Syst. Appl. Microbiol.* **34**(3), 171–179 (2011).
24. Mitrovic, J. et al. The groEL gene as an additional marker for finer differentiation of ‘Candidatus Phytoplasma asteris’-related strains. *Ann. Appl. Biol.* **159**(1), 41–48 (2011).
25. Mitrovic, J. et al. Differentiation of ‘Candidatus Phytoplasma cynodontis’ based on 16S rRNA and groEL genes and identification of a new subgroup, 16Sx14-V. *Plant Dis.* **99**(11), 1578–1583 (2015).
26. Muirhead, K., Pérez-López, E., Bahder, B. W., Hill, J. E. & Dumonceaux, T. The CpnClassifyR is a resource for cpn60 universal target-based classification of phytoplasmas. *Plant Dis.* **103**(10), 2494–2497 (2019).
27. Stämpfli, M., Izadpanah, K., Galetto, L., Salehi, M. & Marzachi, C. Molecular characterization, phylogenetic comparison and serological relationship of the Imp protein of several ‘Candidatus Phytoplasma aurantifolia’ strains. *Plant Pathol.* **62**(2), 452–459 (2013).
28. Danet, J. L. et al. Multilocus sequence analysis reveals the genetic diversity of European fruit tree phytoplasmas and supports the existence of inter-species recombination. *Microbiology* 157(2), 438–450 (2011).

29. Bohunicik, M. et al. Identification of 17 ‘Candidatus Phytoplasma pyri’ genotypes based on the diversity of the imp gene sequence. *Plant Pathol.* 67(4), 971–977 (2018).

30. Marston, E. L., Sumner, J. W. & Regnery, R. L. Evaluation of intraspecies genetic variation within the 60 kDa heat-shock protein gene (groEL) of *Bartonella* species. *Int. J. Syst. Evol. Microbiol.* 49(3), 1015–1023 (1999).

31. Berg, M. et al. Isolation of the gene encoding an immunodominant membrane protein of the apple proliferation phytoplasma, and expression and characterization of the gene product. *Microbiology* 145(8), 1937–1943 (1999).

32. Boonrod, K., Munteanu, B., Jarausch, B., Jarausch, W. & Krczal, G. An immunodominant membrane protein (Imp) of ‘Candidatus Phytoplasma malv’ binds to plant actin. *Mol. Plant-Microbe Interact.* 25(7), 889–895 (2012).

33. Kakizawa, S. et al. Positive selection acting on a surface membrane protein of the plant-pathogenic phytoplasmas. *J. Bacteriol.* 188(9), 3424–3428 (2006).

34. Galetto, L., Rashidi, M., Yamchi, A., Veratti, F. & Marzachi, C. *In vitro* expression of phytoplasma immunodominant membrane proteins. *In: Phytoplasmas and Phytoplasma Diseases Economic Management: How to Reduce Their Economic Impact.* (ed. Bertaccini A) 272–279. (IPWGL International Phytoplasmas Working Group, Bologna, Italy 2014).

35. Trivellone, V. et al. Evidence suggesting interactions between immunodominant membrane protein Imp of Flavescence dorée phytoplasma and protein extracts from distantly related insect species. *J. Appl. Microbiol.* 127(6), 1801–1813 (2019).

36. Rashidi, M. et al. Role of the major antigenic membrane protein in phytoplasma transmission by two insect vector species. *BMC Microbiol.* 15(1), 1–12 (2015).

37. Danet, J. L. et al. Imp and secY, two new markers for MLST (multilocus sequence typing) in the 16SrX phytoplasma taxonomic group. *Bull. Insectol.* 60(2), 339 (2007).

38. Fabre, A. et al. STAMP encoding the antigenic membrane protein of stolbur phytoplasma is useful for molecular epidemiology. *Bull Insectol.* 64, S21–S22 (2011).

39. Arnaud, G. et al. Multilocus sequence typing confirms the close genetic interrelatedness of three distinct flavescence dorée phytoplasma strain clusters and group 16SrV phytoplasmas infecting grapevine and elder in Europe. *Appl. Environ. Microbiol.* 73(12), 4001–4010 (2007).

40. Carraro, L. et al. *Macropsis mendax* as a vector of elm yellows phytoplasma of *Ulmus* species. *Plant Pathol.* 53(1), 90–95 (2004).

41. Nickel, H. The Leafhoppers and Planthoppers of Germany (*Hemiptera, Auchenorrhyncha): Patterns and Strategies in a Highly Diverse Group of Phytophagous Insects* (Pensoft Publishers, New York, 2003).

42. Ahrens, U. & Seemüller, E. Detection of DNA of plant pathogenic mycoplasma-like organisms by a polymerase chain reaction that amplifies a sequence of the 16S rRNA gene. *Phytopathology* 82(8), 828–832 (1993).

43. Caudwell, A., Kuszala, C., Bachelier, J. C. & Larrue, J. Transmission de la *flavescence dorée* de la vigne aux plantes herbacées par l’allongement du temps d’utilisation de la cicadelle *Scaphoideus bittolus* Bail et étude de sa survie sur un grand nombre d’espèces végétales. *Ann. Phytopathol.* 2(2), 415–428 (1970).

44. Thompson, J. D., Gibson, T. J. & Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinform.* 1, 2–3 (2003).

45. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549 (2018).

46. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526 (1993).

47. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426 (1986).

Acknowledgement
This project (no. 22026316) was funded by the “Fachagentur Nachwachsende Rohstoffe e.V.” (FNR), a promoter of the German Federal Ministry for Food and Agriculture.

Author contributions
B.S. performed all experiments and drafted the manuscript. B.H., M.K., C.Z. and B.S. performed sequence data analyses. M.K. designed the study and contributed to manuscript preparation.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-78745-w.

Correspondence and requests for materials should be addressed to B.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
