1. TITLE PAGE

Improved subtyping affords better discrimination of *Trichomonas gallinae* strains and suggests hybrid lineages.

AUTHORS

Abdulwahed F. Alrefaei[1,2] afrefaei@ksu.edu.sa

Richard W. Gerhold[3] rwgerhold@gmail.com

Johanna L Nader[4] J.Nader@uea.ac.uk

Diana J. Bell[1*] D.Bell@uea.ac.uk

Kevin M. Tyler[4] k.tyler@uea.ac.uk

1. University of East Anglia, School of Biological Sciences, Norwich, Norfolk, NR4 7TJ

2. Department of Zoology, King Saud University, College of Science, P.O. Box 2455, Riyadh-11451, Saudi Arabia

3. Department of Biomedical and Diagnostic Sciences, College of Veterinary Medicine, University of Tennessee, Knoxville, TN USA 37996

4. University of East Anglia, School of Medicine, Biomedical Research Centre, Norwich, Norfolk, NR4 7TJ

* To whom correspondence should be addressed.

(Running title.)

Trichomonas gallinae hybrids and subtyping

Abstract

Trichomonas gallinae is a protozoan pathogen that causes avian trichomonosis typically associated with columbids (canker) and birds of prey (frounce) that predate on them, and has recently emerged as an important cause of passerine disease. An archived panel of DNA from North American (USA) birds used initially to establish the ITS ribotypes was reanalysed using Iron hydrogenase (*FeHyd*) gene sequences to provide an alphanumeric subtyping scheme with improved resolution for strain discrimination. Thirteen novel subtypes of *T. gallinae* using *FeHyd* gene as the subtyping locus are described. Although the phylogenetic topologies derived from each single marker are complementary, they are not entirely congruent. This may reflect the complex genetic histories of the isolates analysed which appear to contain two major lineages and several that are hybrid. This new analysis consolidates much of the phylogenetic signal generated from the ITS ribotype and provides additional resolution for discrimination of *T. gallinae* strains. The single copy *FeHyd* gene provides higher resolution genotyping than ITS ribotype alone. It should be used where possible as an additional, single-marker subtyping tool for cultured isolates.

Keywords

Trichomonosis, Frounce, Canker, Subtyping, Ribotyping, Hybrid, Protozoa

Introduction

Natural fauna are affected by a wide range of both macroparasites (such as helminths and arthropods), and micropathogens (such as bacteria, viruses, protozoa, and fungi) (Tompkins et al., 2011). Parasitic diseases can have a significant impact on wild avian species populations by reducing fecundity and nestling survival, and/or by increasing mortality (Newton, 1998). This is particularly true of protozoan parasitic diseases, which include apicomplexan blood infections, such as malaria and leucocytozoonosis and superficial infections such as trichomonosis which can be a major causal factor in nestling mortality and the reduction of adult bird survivorship (Boal et al., 1998; Bunbury et al., 2007, 2008)

Avian trichomonosis has been reported across the world, and in spite of strain specific variation in virulence and the occurrence of parasite carriage without clinical disease, it is considered an important disease for a number of bird species, particularly columbiforms, birds of prey (falconiforms and strigiforms) and more recently passerines (Boal et al., 1998; Bunbury et al., 2008; Lawson et al., 2006; Pennycott et al., 2005; Stabler, 1954). *Trichomonas gallinae* is a causative agent of avian trichomonosis (Forrester and Foster, 2008; Pennycott et al., 2005; Stabler, 1954). This parasite primarily infects the upper digestive tract and can occlude the oesophagus or trachea, leading to eventual death by starvation or suffocation (Stabler, 1954).

Most sequencing investigations used to discriminate *T. gallinae* and other *Trichomonas* spp. lineages have used the internal transcribed spacer regions 1 and 2 (ITS ribotype) (Felleisen, 1997; Gerhold et al., 2008; Grabensteiner et al., 2010; Kleina et al., 2004; Sansano-Maestre et al., 2009). The ITS ribotype is widely used for phylogenetic purposes and is recognised as a robust molecular tool to determine inter- and intra-specific diversity (Amin et al., 2012). Studies have identified over 16 ITS ribotypes in *T. gallinae* species from the USA, Spain and the UK (Anderson et al., 2009; Chi et al., 2013; Gerhold et al., 2008; Grabensteiner et al., 2010; Sansano-Maestre et al., 2009).

The *FeHyd* gene is a second genotyping marker used to differentiate *T. gallinae* isolates (Lawson et al., 2011a). It is a housekeeping gene often used to evaluate evolutionary relationships and which is particularly useful for amitochondrial protists

(Lawson et al., 2011b; Voncken et al., 2002). We previously reported data for the *FeHyd* gene analysed from multiple *T. gallinae* isolates of British origin demonstrated the potential for detecting fine-scale variation between *T. gallinae* strains (Chi et al., 2013).

Here we reanalyse the panel of DNA isolates first used to establish the broad genetic heterogeneity in the *T. gallinae* complex by (Gerhold et al., 2008) based on variation in ITS ribotype to validate Fe hydrogenase as a subtyping locus. Where possible, isolates from the original publication were used: where these were no longer available (having been used up), alternative isolates were used which were matched as closely as possible to the original hosts, geographic location and genotype.

.

Methods

Sources of isolates

Forty-six isolates of *T. gallinae* were obtained from nine different bird species that were collected by Richard Gerhold in the USA between 2005 and 2011. The birds that were collected in the field included seventeen rock pigeons *Columba livia*, nine mourning doves *Zenaida macroura*, six white-winged doves *Zenaida asiatica*, four band-tailed pigeons *Patagioenas fasciata*, four Eurasian collared doves *Streptopelia decaocto*, two common ground-doves *Columbina passerina*, two Cooper's hawks *Accipiter cooperii*, one house finch *Haemorhous mexicanus*, and one ring-necked dove *Streptopelia risoria*. Three additional isolates were obtained from the American Type Culture Collection (ATCC). Samples were obtained from different states of the USA (Arizona, California, Colorado, Georgia, Kentucky, Massachusetts, Tennessee, Texas and Virginia) (Fig. 1; Table 1). Although ITS ribotype was previously obtained for many of the isolates (Gerhold et al., 2008), all samples tabulated in Table 1 were re-sequenced as part of this study.

DNA from cultured isolates.

DNA was extracted from the 46 cultured isolates (Table 1) as previous described (Gerhold et al., 2008).

PCR of the Fe-hydrogenase gene

Polymerase chain reaction (PCR) was used to amplify fragments of the *FeHyd* gene using the primers TrichhydFOR (5'-GTTTGGGATGGCCTCAGAAT-3') and TrichhydREV (5'-AGCCGAAGATGTTGTCGAAT-3') as described by Lawson et al. (2011). The PCR mix included 3 µL of DNA in a 47 µL reaction containing 10 µL of 5X buffer (Qiagen, UK), 3 µL of 25 mM MgCl2 (Qiagen, UK), 0.4 µL 10 mM dNTP mix (Qiagen, UK), 2.5 µL each of 5 µM forward and reverse primers solutions (Eurofins Genomics, Germany), 0.25µL of 5 U/µM HotStarTaq *Plus* DNA polymerase (Qiagen, UK), and 28.35 µL of nuclease-free water (Promega UK Ltd, Southampton, UK). Each

PCR run contained a negative control of nuclease-free water and positive controls of previously extracted *T. gallinae* DNA from a British greenfinch *Chloris chloris* obtained in 2007. The PCR amplification was performed using the following cycling conditions: 94 °C for 15 minutes, followed by 35 cycles of 94 °C for 1 minute and 52 °C for 30 seconds, then 72 °C for 1 minute, and a final extension at 72 °C for 5 min. PCR amplification was confirmed visually under ultraviolet light by using a 1% agarose gel, stained with ethidium bromide, and the expected product size was approximately 1 kb. The PCR products were submitted for sequencing with both TrichhydFOR and TrichhydREV primers to Source BioScience (Nottingham, UK).

Sequence analysis and phylogenetic trees

The molecular and phylogenetic relationships of the sequences obtained for *T. gallinae* were determined using Molecular Evolutionary Genetics Analysis (MEGA) software version 6.06 (Tamura et al., 2011). In order to investigate their relatedness to other trichomonad isolates, phylogenetic analysis of these sequences was performed. Chromatograph files were inspected and refined using the MEGA 6.06 Trace Data File Viewer/Editor extension. All sequence data were aligned using sequence from the forward primers and reverse complement of sequence from the reverse primers. Existing *T. gallinae* sequences used as comparators in the study were obtained from the National Center for Biotechnology Information (NCBI) GenBank database (Table 3.2). Cladogram trees of the datasets obtained from the *FeHyd* sequences were constructed separately using the maximum likelihood (ML) method and the Tamura-Nei model to analyse taxa relationships by nucleotide analysis (Saitou and Nei, 1987; Tamura et al., 2011). Felsenstein's bootstrap test was used to calculate associated taxa clustered in the bootstrap values (2,000 times) (Felsenstein, 1985).

Consensus Network Analysis

SplitsTree version 4.14.3 (Huson and Bryant, 2006) was used to construct a NeighborNet for the ITS ribotype, *FeHyd* gene fragment and the concatenated

sequence of the *FeHyd* gene fragment and the ITS ribotype after which a final consensus network, obtained from the combination of the individual ML tree for each gene, was constructed.

Analysis of Fe-hydrogenase Sequences

Levels of genetic differentiation, gene flow, and allelic diversity among Fe-hydrogenase sequences were assessed using the DnaSP program v. 5.10.1 (Librado and Rozas, 2009). Tajima's D (Tajima, 1989) and Fu and Li's D (Fu and Li, 1993) statistics were used to detect signals of selection and gene flow within defined populations of four or more sequences, estimating values as moderately significant ($P < 0.05$), significant ($P < 0.01$), or highly significant ($P < 0.001$). Sequence diversity (π) and allelic diversity were determined using Nei's method (Nei and Kumar, 2000) for populations of two or more sequences. Genetic differentiation was described using F_{st} (Fu and Li, 1993), providing a model of population structure that defined differentiation as negligible ($F_{st} < 0.15$), moderate ($0.15 < F_{st} < 0.25$), and high ($F_{st} > 0.25$) (Wright, 1978). Absolute diversity between allele populations was further assessed using D_{xy}, according to Nei's method.

Results

Phylogenetic analysis of ITS ribotpe sequencing

Table 1 details the ITS ribotypes for the panel of isolates. ITS ribotype amplicon sequences were plotted phylogenetically as a maximum likelihood (ML) tree in Fig 2A with the *T. vaginalis* ITS ribotype being used to provide an outgroup. According to these results, the isolates of *T. gallinae* in USA were divided in four well supported groups (A,B),(C,D,E),(F,G),(H,I,J,K,L). With clear support for two major lineages (A,B,C,D) and (H,I,J,K,L) and with isolates with F and G forming a third sub-clade in between the two larger lineage-based clades, thereby indicating that these isolates may belong to an intermediate ribotype. These results were essentially identical to the findings of the previous study (Gerhold et al., 2008).

Phylogenetic analysis of *FeHyd* gene sequencing

Sequence comparisons of *FeHyd* amplicons obtained from the 46 isolates of *T. gallinae* were used to provide subtypes for the isolates (Table 1). Thirteen new subtypes were identified and these were used in conjunction with 10 previously subtyped (Table 2) isolates for subsequent phylogenetic analysis. Unfortunately, lack of sufficient sample precluded sequencing of a ribotype B FeHyd amplicon. An initial phylogenetic analysis using a *T. vaginalis* ortholog of *FeHyd* gene as an outgroup provided support for four groups, resolving the ribotype K and ribotype L isolates into discrete lineages as well as supporting ribotypes C, D and E and ribotype A as distinct lineages (Fig. 2B). Intriguingly, although the *FeHyd* gene nucleotide sequence provides considerable additional discriminatory power for strains of most ribotypes, the *FeHyd* sequencing of ribotypes G and H isolates failed even to support these isolates as a discrete group from ribotype A by use of this locus alone. It was also noted that isolates recently described as *Trichomonas stableri*, which has a K ribotype, possessed *FeHyd* gene sequences which were most akin to the ribotype L isolates. It is notable that in this study, 15 isolates of *T. gallinae* were identical to subtype A1 (GenBank JF681136) which was isolated from a UK finch and described by Lawson et al. (2011a). In addition, seven isolates were 100% identical to the subtype A2 sequence JF681141 that was obtained from a Madagascar turtle dove *Streptopelia*

picturata on Mahé, an island in the Seychelles (Lawson et al., 2011a). These genotypes appeared to be widespread globally and stable. By comparison, five new L subtypes were resolved, many of these from the same host species in the same year over a limited geographic range suggesting that the *FeHyd* nucleotide sequence for these strains is far more polymorphic.

Phylogenetic analysis of the concatenated sequences

The analysis of concatenated sequences can improve the quality of the phylogenetic reconstruction and optimizes the taxonomic resolution, reducing stochastic effects associated with a short sequence length. The nucleotide sequence data for the two loci concatenated produced a sequence length of 991bp. The ML tree generated resolved six well supported groups, namely 1) isolates of ribotype A, 2) isolates of ribotypes C,D,E with some evidence that C4 is divergent from other isolates in this group, 3) isolates of ribotype F and G, 4) isolates of ribotype H, 5) isolates of ribotype K and 6) isolates of ribotype L (Fig 2C). Because of the greater variation in the *FeHyd* gene, although the *T. stableri* isolates were ribotype K, they actually group with the ribotype L isolates on a concatenated tree. Although the topologies of the trees obtained are incongruent for a small number of phyla, from a typing perspective the two loci introduce more discriminating characters, enabling greater resolution than either locus alone

Splits Trees analysis

Phylogenetic analyses aim to detect the evolutionary relationships between different species or taxa, in order to understand the evolutionary distances and relative positioning of strains. Phylogenetic trees are widely used to address this task and are usually computed from molecular sequences. However, the use of trees is a less suited method to model mechanisms of reticulate evolution (Sneath 1975), such as recombination, horizontal gene transfer, or hybridization. Phylogenetic networks such as splits network analyses supply an alternative to phylogenetic trees when analysing data sets whose evolution involves significant amounts of reticulate events (Griffiths

and Marjoram 1996; Rieseberg 1997; Doolittle 1999). Furthermore, even for a set of taxa that have evolved according to a tree-based model of evolution, phylogenetic networks can be usefully employed to explicitly represent conflicts in a data set that may be caused by mechanisms such as incomplete lineage sorting or by the inadequacies of an assumed evolutionary model (Huson and Bryant 2006). When hybridization manifests as phylogenetic incongruence between different genes, hybridization may be visualized in networks (Huson and Bryant, 2006). Since the ITS and *FeHyd* gene ML trees were at variance in how they clustered some isolates, the phylogenetic signal was analysed in more detail using split networks to consider if they could provide evidence for gene flow, hybridization or homoplasy.

Neighbour network analysis for the ITS ribotype sequence represented by the splits tree (Fig.3A) grouped isolates in agreement with the earlier phylogenetic tree (Fig. 2A). It is clear that ribotypes (A-E), (H-L) and (F and G) all represent distinct groups. There is some support for hybridization at this locus within the A-E and H-L groups. The positioning of the ribotype F and ribotype G strains does suggest that this group may represent a hybrid lineage formed from the other two groups. The central positioning of strains F and G on the ITS split network tree, directly in between the two larger clades, indicates that it shares approximately half of its sequence data with either of the other lineages, thereby indicating the potential that these strains represent hybrid isolates.

Neighbour network analysis for the FeHyd gene sequence (Fig 3B) resulted in a richer network than for the ITS ribotype due to the higher degree of character variation. Topologically, genotypes A1-H1 associate as a large group of very similar phylogenetically discrete alleles from the genotype L group, which has considerable variation in its *FeHyd* sequences. The wide variety of paths between these groups and the very well supported ribotype K group suggesting that at the *FeHyd* locus, the ribotype K isolates may be hybrid. This same trend was identified for strain K on the *FeHyd* split network tree, again presenting the possibility that this strain is positioned in the intermediate way because it represents a hybrid strain between the two larger lineages.

Neighbour network analysis for the concatenated sequences from the ITS region and *FeHyd* gene highlights the phylogenetic distance between the two major lineages.

Although the larger number of discriminant characters derived from *FeHyd* mean that the concatenated network has broadly similar topology to that from the *FeHyd* gene alone, the addition of the ITS region (Fig. 3C), serves to emphasize that ribotypes F, G, H, K are likely to be hybrids (at least at these two loci) between these two independently evolving major lineages.

Allelic diversity and genetic differentiation of Fe-hydrogenase

Fe-hydrogenase sequences showed a uniformly high allelic diversity across population groups (mean: 0.934, range: 0.667-1.000), and relatively high and variable nucleotide diversities (mean π: 0.039, 95% CI: 0.007 – 0.071) (Table 3). *T. gallinae* isolates from white-winged doves in the USA exhibited the highest inter-allelic nucleotide diversity at the Fe-hydrogenase locus (π = 0.090), while isolates from common ground doves in the USA showed the lowest intra-allelic nucleotide diversity by more than ten-fold (π = 0.008), but the two groups did not differ when evaluated for selective pressure (dN/dS = 0.102 and 0.108, respectively; data not shown). No deviation from the expected model for neutral evolution could be detected within populations, with non-significant Tajima's D and Fu and Li's D values ($P > 0.10$) across all comparisons for which sufficient sequence data was available.

Most of the sampled Fe-hydrogenase allele populations were highly genetically differentiated (mean F_{st} = 0.373, 95% CI: 0.231-0.515) and diverged (mean D_{xy} = 0.070, 95% CI: 0.052-0.087), although there were slight variations between and within populations of *T. gallinae* from the United Kingdom versus the North America (Table 4). 10 out of 15 (75%) allele populations were highly differentiated ($F_{st} > 0.25$) at the Fe-hydrogenase locus, while band-tailed pigeon *T. gallinae* populations were only moderately differentiated from North American alleles as a whole (F_{st} = 0.214), and the North American common ground dove was only marginally differentiated from UK-based populations (F_{st} = 0.15603 and 0.16092). Wood pigeons from the United Kingdom, on the other hand, showed an absence of genetic differentiation when compared to the entirety of UK Fe-hydrogenase sequences (F_{st} = -0.212), which complements the lowest overall degree of absolute divergence (D_{xy} = 0.012) observed for this comparison.

Discussion

In 1938, Robert Stabler published a discussion in which he refuted the usage of the name *T. columbae* for trichomonad parasites causing canker in the mouth, crop and liver of pigeons asserting *gallinae* as the species name with priority having been first described by the celebrated Italian veterinary microbiologist Sebastiano Rivolta (Rivolta, 1878). In subsequent publication, he provided detailed description of a parasite displaying extreme pleomorphism, which was morphologically indistinguishable from the human pathogen *T. vaginalis* (Stabler, 1938). Since then, many investigators have noted the difficulty of accommodating a group of parasites with such variable morphology, promiscuity in host range and wide differential in ability to cause disease within a single species taxon.

Early molecular work on virulence showed clear differences between strains in their antigenic and isoenzyme repertoire (Nadler and Honigberg, 1988). More recently, as better genetic tools have become widely available, it has become even clearer that trichomonads which infect the crops of columbids are far from homogenous, but are actually varied substantially in their genetics (Gerhold et al., 2008; Anderson et al., 2009; Lawson et al., 2011; Girard et al., 2014). In 2008, the genetic diversity of parasites infecting the crop of columbids in the USA was systematically evaluated using the ITS ribotype for 42 diverse isolates establishing 12 distinct ribotypes, and several additional ribotypes have since been discovered (Gerhold et al., 2008). In that paper it was also noted that the variation at the sequenced locus was so great that *T. vaginalis* and several other species of trichomonas could be accommodated within the diversity observed (Gerhold et al., 2008). Indeed, the *T. vaginalis* ITS ribotype had a closer relevance to some of the ribotypes (HIJKL) than to those ribotypes typical for the other major branches (ABCDE) (Gerhold et al., 2008). Despite the close overall genomic relationship between *T. gallinae* and *T. vaginalis*, sufficient nucleotide diversity exists at the markers used for strain discrimination in this study for *T. vaginalis* to be used as an effective outgroup.

The essence of species recognition for eukaryotic pathogens is the matching of genetic difference, reproductive segregation and significant differences in phenotype - be they morphological, pathological or related to host specificity. The advent of ribotyping has enabled investigators to consider whether some ribotypes that are

genetically different, may be phenotypically different too. The suggestion has been made that where ribotype A isolates are normally virulent, ribotype C isolates are not - although this has not yet led to formal suggestion of speciation. More recently it was observed that isolates from band-tailed pigeons of ribotype K appeared to be smaller than some other *T. gallinae* ribotypes (Girard et al., 2014). The authors sequenced the isolates at three loci to illustrate that this group were genetically distinct and suggested that they be referred to by a new species name *T. stableri.*

Our data of this study based on comparison of a wider set of isolates at two loci suggests that the genetic definition provided for *T. stableri* is premature. It appears increasingly likely that the complex of lineages encompassed in this study contain hybrid lineages. Our data of this study are consistent with ribotype L and ribotypes A,B,C,D,E represent two independently evolving metapopulations. If genomes provide confirmation of this and there is phenotypic support, new species classification might be warranted (Fernando et al., 2016). Hybrid lineages such as F, G, H, K, and "*T. stableri*" (and possibly *T. vaginalis*) confound simple taxonomic rules. A similar situation exists currently for *Trypanosoma cruzi* with two parental lineages and multiple hybrid lineages where these are known as disease typing units (DTUs) (Westenberger et al., 2005). For avian trichomonosis there is currently no clinical imperative to discriminate between the trichomonads causing canker in the mouth, crop and liver of columbids (Chi et al., 2013). The lesions and therapeutic options are ostensibly the same and so it seems prudent to await consideration of representative genomes from the major lineages and multilocus-sequence typing across a broad span of isolates before initiating taxonomic change which may sew confusion amongst clinicians. In the interim, the use of ITS ribotype and FeHyd gene sequencing presented here represents a validated, quick and easy alphanumeric typing scheme which is hoped to be helpful to the community for strain discrimination (Gerhold et al., 2008; Lawson et al., 2011a; Chi et al., 2013; Girard et al., 2014).

Genetic differentiation and non-neutral selection analyses of Fe-hydrogenase sequences in this study provide an important insight into population structure and diversity at this locus in geographically (and phenotypically) diverse *T. gallinae* isolates. Allelic diversity and nucleotide diversity were high within all defined populations, but the overwhelming predominance of synonymous versus non-

synonymous mutations (i.e. significant negative selection; data not shown) indicates that mutations are likely the result of non-adaptive evolution, and that an incentive exists for *T. gallinae* to preserve the biological function of Fe-hydrogenase, irrespective of geographic origin or host type. This is mirrored by the results from both neutrality tests, for which no significant deviation from neutral evolution could be identified. The excess of random, neutral mutations which have accumulated at this locus appears to be unique to individual alleles, and have resulted in strongly differentiated populations, with a majority of comparisons revealing strong population structures and an absence of any significant gene flow or panmixis of global alleles (F_{st} > 0.25). The North American *T. gallinae* isolates from common ground doves deviated from this trend, however, exhibiting considerably less divergence and differentiation when compared to UK alleles than to other North American populations (D_{xy} = 0.156 versus 0.342, respectively). This complements the phylogenetic proximity of these isolates at this locus as previously shown (Figure 2B). The closer genetic relationship between these North American Fe-hydrogenase alleles to UK alleles is a possible indication of convergent evolution at this locus, or potentially the consequence of a hybridization event, but the limited sample size and single genetic target used in this study makes it difficult to confirm or refute either. Future studies that expand the analyses to include whole genome sequences, and a larger and more epidemiologically-diverse dataset, are important in determining whether these discrepancies are unique to Fe-hydrogenase, or reflected by genome-wide trends, and will help to unfold the diversity and population structure of global *T. gallinae* isolates on a larger scale.

Data Availability: All new sequences generated are available through GenBank (https://www.ncbi.nlm.nih.gov/genbank/) with the accession numbers provided.

Competing Interests: The authors declare they have no competing interests.

Funding: AFA's studentship at the University of East Anglia was funded by the King Saud University, Saudi Arabia.

Author contributions. KMT and RG conceived the study. RG provided the samples for analysis. AFA performed the laboratory work. AFA, JLN and KMT undertook the phylogenetic analyses. AFA, KMT and DJB drafted the paper. All authors refined the draft and approved it prior to submission.

Ethics: No local ethical committee or animal ethics committee or permission to carry out fieldwork was required for this set of studies.

Acknowledgements: We thank Graham Clark and Dietmar Steverding for critical reading of the dissertation from which this manuscript is drawn.

Figure Legends:

Fig. 1. Geographical distribution of the sites in the United States where *Trichomonas gallinae* isolates were collected between 2005 and 2011. Arizona, 4 birds; California, 4 birds; Colorado, 1 bird; Georgia, 15 birds; Kentucky, 2 birds; Massachusetts, 1 bird; Tennessee, 3 birds; Texas, 11 birds; Virginia, 3 birds.

Figure 2 Phylogenetic tree based analysis for *Trichomonas gallinae*. 2A 5.8S rRNA gene and flanking ITS ribotypes using 337-bp alignment. 2B. *FeHyd* gene regions in samples isolated from the birds described in Table 1 using 653-bp alignment. 2C Concatenated 991bp sequence alignment. The bootstrap trees followed the Tamura–Nei model and used and 2,000 replicates. Tree topology was tested using the maximum likelihood method. Bootstrap values less than 50% are not shown. Letters and numerals correspond to the previously described *FeHyd* gene sequences of *T. gallinae*. The black dots indicate the new genotypes reported in this study.

Figure 3 Splits network analysis of *Trichomonas gallinae*. 3A Network based on ITS region sequences. 3B FeHyd sequences. 3C Concatenated 991bp sequences. The consensus phylogenetic network (neighbour-net) dendrograms were calculated using SplitsTree with an equal angle algorithm.

Table 1 Panel of genomic DNA from cultured isolates of *Trichomonas gallinae* subtyped as part of this study.

Isolate ID	Case no.	Species	Year found	Location	Lesions present	ITS type	Fe-hydrogenase subtype
1	M2	Mourning dove	2004	Georgia	Yes	A	A1
2	M6	Mourning dove	2004	Arizona	Yes	A	A1
3	M8	Mourning dove	2004	Georgia	Yes	L	L4
4	M11	Mourning dove	2004	Georgia	No	A	A1
5	M21	Mourning dove	2006	Kentucky	Yes	A	A1
6	WW840	White-winged dove	2006	Texas	No	H	H1
7	WW1208	White-winged dove	2006	Texas	No	L	L2
8	WW1200	White-winged dove	2006	Texas	No	L	L5
9	R11	Rock pigeon	2004	Georgia	No	A	A1
10	R19	Rock pigeon	2004	Georgia	No	A	A2
11	R14	Rock pigeon	2004	Georgia	No	A	A1
12	R22	Rock pigeon	2004	Georgia	No	A	A2
13	RN3	Ring-necked dove*	2006	Tennessee	No	D	D1
14	R11	Rock pigeon	2004	Georgia	No	A	A2
15	R15	Rock pigeon	2004	Georgia	No	A	A2
16	R32	Rock pigeon	2004	Georgia	No	A	A1
17	BTPN1	Band tailed pigeon	2007	California	Yes	A	A2
18	GD1	Common ground dove	2006	Texas	No	G	G1
19	TG	Rock pigeon	1947	Unknown	No	C	C1
20	COHA1	Cooper's hawk	2004	Arizona	Yes	L	L1
21	BTPN4	Band-tailed pigeon	2007	California	Yes	K	K1.1
22	ECD829	Eurasian collared-dove	2006	Texas	No	D	D1
23	RODO1619	Rock pigeon	2011	Tennessee	unknown	A	A1
24	SGC	Rock pigeon	1968	Colorado	No	C	C4
25	VA2	Mourning dove	2006	Virginia	Yes	A	A1
26	M9	Mourning dove	2004	Arizona	Yes	L	L4

27	GD1321	Common ground dove	2006	Texas	No	F	F1
28	R11	Rock pigeon	2004	Georgia	No	A	A2
29	R51	Rock pigeon	2006	Georgia	No	A	A2
30	RD1620	Rock pigeon	2011	Tennessee	No	A	A1
31	EDC858	Eurasian collared-dove	2006	Texas	No	E	E1
32	WW947	White-winged dove	2006	Texas	No	L	L1
33	CC298-07	Mourning dove	2007	Virginia	Yes	A	A1
34	C3	Cooper's hawk	2004	Arizona	Yes	L	L4
35	HF1	House finch	2006	Kentucky	Yes	A	A1
36	WW1159	White-winged dove	2006	Texas	No	I	dna
37	R14	Rock pigeon	2004	Georgia	No	A	A1
38	AG	Rock pigeon	1956	Massachusetts	unknown	A	A1
39	ECD1632	Eurasian collared-dove	unknown	unknown	unknown	A	A1
40	VA1	Mourning dove	2007	Virginia	Yes	A	A1
41	R32	Rock pigeon	2004	Georgia	No	A	A1
42	WW1323	White-winged dove	2006	Texas	No	L	L3
43	BTPN3	Band-tailed pigeon	2007	California	Yes	K	K2
44	BTPN2	Band-tailed pigeon	2007	California	Yes	K	K1
45	R28	Rock pigeon	2004	Georgia	No	D	D1
46	ECD1053	Eurasian collared-dove	2006	Texas	No	D	D1

*All DNA was from wild birds, except isolate 13, which was a captive, ring-necked dove. DNA – did not amplify.

Table 2 Isolates of *Trichomonas gallinae* and their subtypes used for phylogenetic analysis

Species (host)	Genotype	Origin	GenBank	Reference
T. gallinae (Greenfinch *Chloris chloris*)	A1	UK	JF681136	(Lawson et al., 2011a)
T. gallinae (Sparrowhawk *Accipiter nisus*)	A1.1	UK	KC529660	(Chi et al., 2013)
T. gallinae (Wood pigeon *Columba palumbus*)	A1.2	UK	KC962158	(Chi et al., 2013)
T. gallinae (Wood pigeon *Columba palumbus*)	A1.3	UK	KC529661	(Chi et al., 2013)
T. gallinae (Madagascar turtle dove *Nesoenas picturatus*)	A2	Seychelles	JF681141	(Lawson et al., 2011a)
T. gallinae ATCC 30230 (Rock pigeon *Columba palumbus*)	C1	North America	Identical to AF446077	(Lawson et al., 2011a)
T. gallinae (Wood pigeon *Columba palumbus*)	C2	UK	KC529664	(Chi et al., 2013)
T. gallinae (Collared dove *Columba livia*)	C3	UK	KC529663	(Chi et al., 2013)
T. gallinae (Wood pigeon *Columba palumbus*)	C4	UK	KC529662	(Chi et al., 2013)
T. gallinae (Rock pigeon *Columba livia*)	D1	USA	KY496778	This study
T. gallinae (Eurasian collared-dove *Streptopelia decaocto*)	E1	USA	KY496781	This study
T. gallinae (Common ground dove)	F1	USA	KY496782	This study
T. gallinae (Common ground dove)	G1	USA	KY496783	This study
T. gallinae (White-winged dove)	H1	USA	KY496784	This study
T. gallinae (Band-tailed pigeon *Patagioenas fasciata*)	K1	USA	KY496786	This study
T. gallinae (Band-tailed pigeon *Patagioenas fasciata*)	K1.1	USA	KY496787	This study
T. gallinae (Band-tailed pigeon *Patagioenas fasciata*)	K2	USA	KY496785	This study
T. gallinae (White-winged dove *Zenaida asiatica*)	L1	USA	KY496788	This study
T. gallinae (White-winged dove *Zenaida asiatica*)	L2	USA	KY496791	This study
T. gallinae (White-winged dove *Zenaida asiatica*)	L3	USA	KY496790	This study
T. gallinae (Mourning dove *Zenaida macroura*)	L4	USA	KY496789	This study
T. gallinae (White-winged dove *Zenaida asiatica*)	L5	USA	KY496792	This study
T. stableri (Band-tailed pigeons *Patagioenas fasciata*)	K3	USA	KC660123	(Girard et al., 2014)
T. vaginalis G3 (Human)	NA	UK	XM_001310179	(Carlton et al., 2007)

Table 3 Fe-hydrogenase allelic diversity and signals of non-neutral selection for geographically-diverse *T. gallinae* isolates from various avian host species

Origin	Host Species	Sequences	Alleles	Allelic Diversity	Pi (π)	Tajima's D	Fu & Li's D
UK	All	7	6	0.95238	0.01408	-0.01034	-0.26925
UK	*Columba palumbus* (Wood pigeon)	4	4	1.00000	0.01567	0.70444	0.70444
USA	All	13	12	0.98718	0.08967	1.29078	0.95873
USA	*Columbina passerina* (Common ground dove)	2	2	1.00000	0.00768	‡	‡
USA	*Patagioenas fasciata* (Band-tailed pigeon)	3	2	0.66667	0.01919	‡	‡
USA	*Zenaida asiatica* (White-winged dove)	5	5	1.00000	0.09002	-0.24770	-0.23202

‡ Does not meet minimum 4 sequence requirement for tests of neutrality (Tajima's D and Fu & Li's D)
* Sig. $P < 0.05$

Table 4. Fe-hydrogenase genetic differentiation estimates for geographically-diverse *T. gallinae* isolates from various avian hosts: F_{st} values (bottom left) and D_{xy} (top right)

	United Kingdom (All)	United Kingdom (*Columba palumbus*)	United States (All)	United States (*Columbina passerina*)	United States (*Patagioenas fasciata*)	United States (*Zenaida asiatica*)
United Kingdom (All)		0.01227	0.07500	0.01289	0.06087	0.11456
United Kingdom (*Columba palumbus*)	-0.21229		0.07508	0.01392	0.06094	0.11459
USA (All)	0.30840*	0.29843*		0.07404	0.06925	0.09139
USA (*Columbina passerina*)	0.15603*	0.16092*	0.34264*		0.0595	0.11324
USA (*Patagioenas fasciata*)	0.72673*	0.71391*	0.21393*	0.77419*		0.09738
USA (*Zenaida asiatica*)	0.54568*	0.53881*	0.01694	0.56864*	0.43922*	

* F_{st} Sig. $P < 0.05$

References:

Amin, A., Bilic, I., Berger, E., Hess, M., 2012. Trichomonas gallinae, in comparison to Tetratrichomonas gallinarum, induces distinctive cytopathogenic effects in tissue cultures. Veterinary parasitology 186, 196-206.

Anderson, N.L., Grahn, R.A., Van Hoosear, K., BonDurant, R.H., 2009. Studies of trichomonad protozoa in free ranging songbirds: prevalence of Trichomonas gallinae in house finches (Carpodacus mexicanus) and corvids and a novel trichomonad in mockingbirds (Mimus polyglottos). Veterinary parasitology 161, 178-186.

Boal, C.W., Mannan, R.W., Hudelson, K.S., 1998. Trichomoniasis in Cooper's hawks from Arizona. Journal of Wildlife Diseases 34, 590-593.

Bunbury, N., Jones, C., Greenwood, A., Bell, D., 2007. Trichomonas gallinae in Mauritian columbids: implications for an endangered endemic. Journal of Wildlife Diseases 43, 399-407.

Bunbury, N., Jones, C., Greenwood, A., Bell, D., 2008. Epidemiology and conservation implications of Trichomonas gallinae infection in the endangered Mauritian pink pigeon. Biological Conservation 141, 153-161.

Carlton, J.M., Hirt, R.P., Silva, J.C., Delcher, A.L., Schatz, M., Zhao, Q., Wortman, J.R., Bidwell, S.L., Alsmark, U.C.M., Besteiro, S., 2007. Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. Science 315, 207-212.

Chi, J.F., Lawson, B., Durrant, C., Beckmann, K., John, S., Alrefaei, A.F., Kirkbride, K., Bell, D.J., Cunningham, A.A., Tyler, K.M., 2013. The finch epidemic strain of Trichomonas gallinae is predominant in British non-passerines. Parasitology 140, 1234-1245.

Felleisen, R., 1997. Comparative sequence analysis of 5· 8S rRNA genes and internal transcribed spacer (ITS) regions of trichomonadid protozoa. Parasitology 115, 111-119.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution, 783-791.

Forrester, D.J., Foster, G.W., 2008. Trichomonosis. Parasitic diseases of wild birds, 120-153.

Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693-709.

Gerhold, R.W., Yabsley, M.J., Smith, A.J., Ostergaard, E., Mannan, W., Cann, J.D., Fischer, J.R., 2008. Molecular characterization of the Trichomonas gallinae morphologic complex in the United States. Journal of Parasitology 94, 1335-1341.

Girard, Y.A., Rogers, K.H., Gerhold, R., Land, K.M., Lenaghan, S.C., Woods, L.W., Haberkern, N., Hopper, M., Cann, J.D., Johnson, C.K., 2014. Trichomonas stableri n. sp., an agent of trichomonosis in Pacific Coast band-tailed pigeons (Patagioenas fasciata monilis). International Journal for Parasitology: Parasites and Wildlife 3, 32-40.

Grabensteiner, E., Bilic, I., Kolbe, T., Hess, M., 2010. Molecular analysis of clonal trichomonad isolates indicate the existence of heterogenic species present in different birds and within the same host. Veterinary Parasitology 172, 53-64.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution 23, 254-267.

Kleina, P., Bettim-Bandinelli, J., Bonatto, S.L.s., Benchimol, M., Bogo, M.c.R., 2004. Molecular phylogeny of Trichomonadidae family inferred from ITS-1, 5.8 S rRNA and ITS-2 sequences. International journal for parasitology 34, 963-970.

Lawson, B., Cunningham, A., Chantrey, J., Hughes, L., Kirkwood, J., Pennycott, T., Simpson, V., 2006. Epidemic finch mortality. Veterinary Record 159, 367-367.

Lawson, B., Cunningham, A.A., Chantrey, J., Hughes, L.A., John, S.K., Bunbury, N., Bell, D.J., Tyler, K.M., 2011a. A clonal strain of Trichomonas gallinae is the aetiologic agent of an emerging avian epidemic disease. Infection, Genetics and Evolution 11, 1638-1645.

Lawson, B., Robinson, R.A., Neimanis, A., Handeland, K., Isomursu, M., Agren, E.O., Hamnes, I.S., Tyler, K.M., Chantrey, J., Hughes, L.A., 2011b. Evidence of spread of the emerging infectious disease, finch trichomonosis, by migrating birds. EcoHealth 8, 143-153.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451-1452.

Nadler, S.A., Honigberg, B., 1988. Genetic differentiation and biochemical polymorphism among trichomonads. The Journal of parasitology, 797-804.

Nei, M., Kumar, S., 2000. Molecular evolution and phylogenetics. Oxford university press.

Newton, I., 1998. Population limitation in birds. Academic press.

Pennycott, T., Lawson, B., Cunningham, A., Simpson, V., Chantrey, J., 2005. Necrotic ingluvitis in wild finches. Veterinary Record 157, 360-360.

Rivolta, S., 1878. Una forma di croup prodotta da un infusorio, nei polli. Giorn. anat., fisiol. e patol. anim 10, 149-158.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution 4, 406-425.
Sansano-Maestre, J., Garijo-Toledo, M.M., Gómez-Muñoz, M.T., 2009. Prevalence and genotyping of Trichomonas gallinae in pigeons and birds of prey. Avian Pathology 38, 201-207.
Stabler, R.M., 1954. Trichomonas gallinae: a review. Experimental Parasitology 3, 368-402.
Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585-595.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution 28, 2731-2739.
Tompkins, D.M., Dunn, A.M., Smith, M.J., Telfer, S., 2011. Wildlife diseases: from individuals to ecosystems. Journal of Animal Ecology 80, 19-38.
Voncken, F., Boxma, B., Tjaden, J., Akhmanova, A., Huynen, M., Verbeek, F., Tielens, A.G., Haferkamp, I., Neuhaus, H.E., Vogels, G., 2002. Multiple origins of hydrogenosomes: functional and phylogenetic evidence from the ADP/ATP carrier of the anaerobic chytrid Neocallimastix sp. Molecular microbiology 44, 1441-1454.
Westenberger, S.J., Barnabe, C., Campbell, D.A., Sturm, N.R., 2005. Two hybridization events define the population structure of Trypanosoma cruzi. Genetics 171, 527-543.
Wright, S., 1978. Evolution and the Gentics of Populations. Variability within and among Natural Population. . University of Chicago Press, Chicago, Illinois, USA.

Fig 1

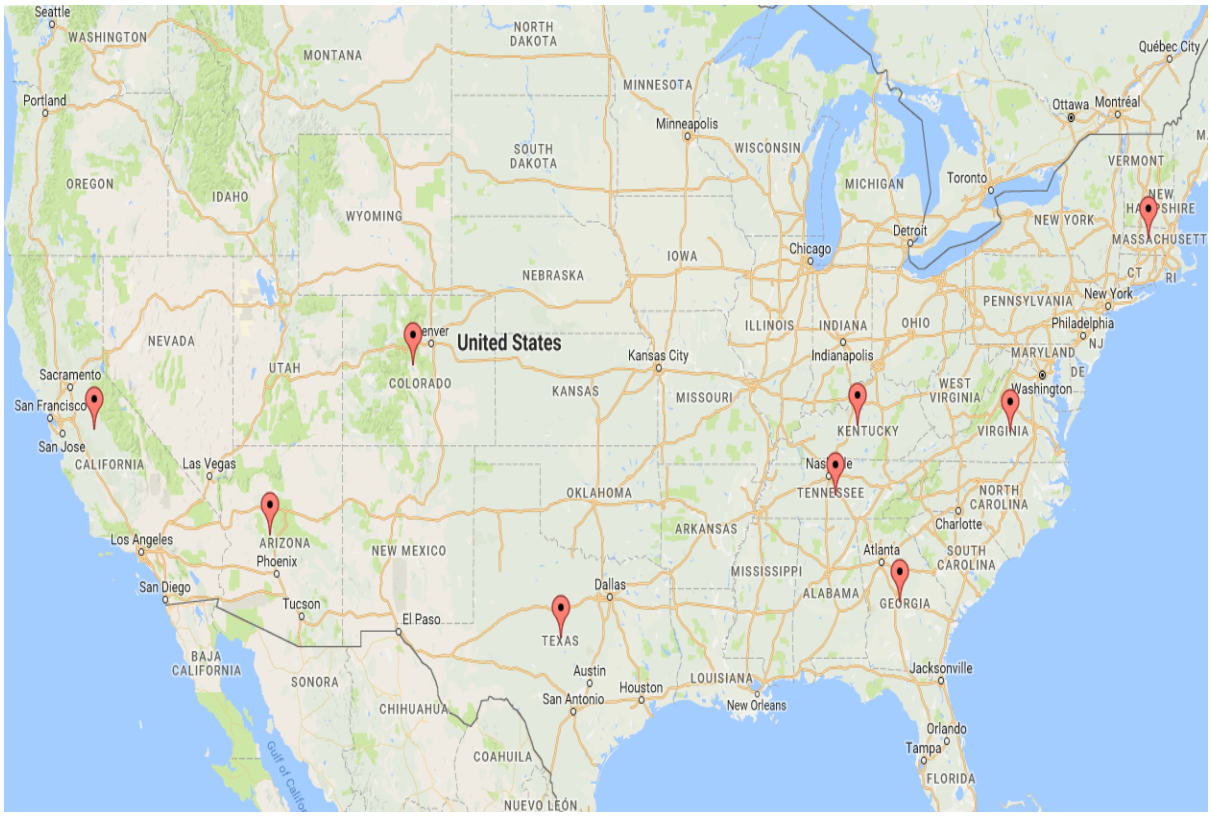
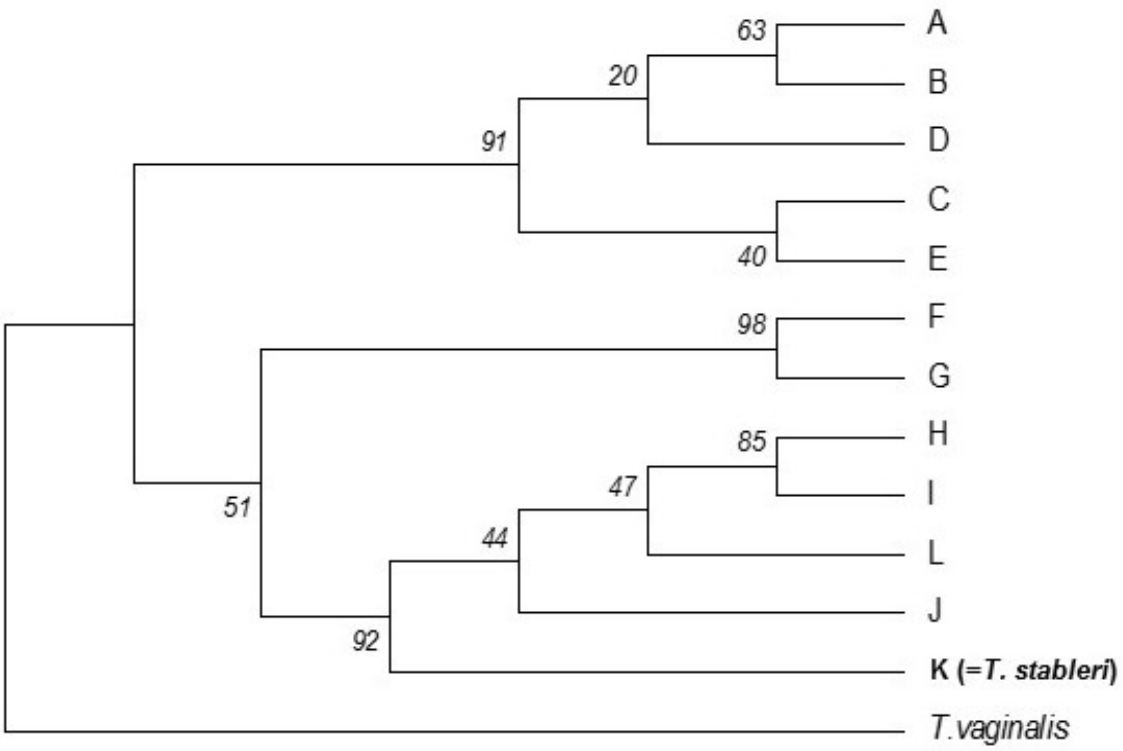
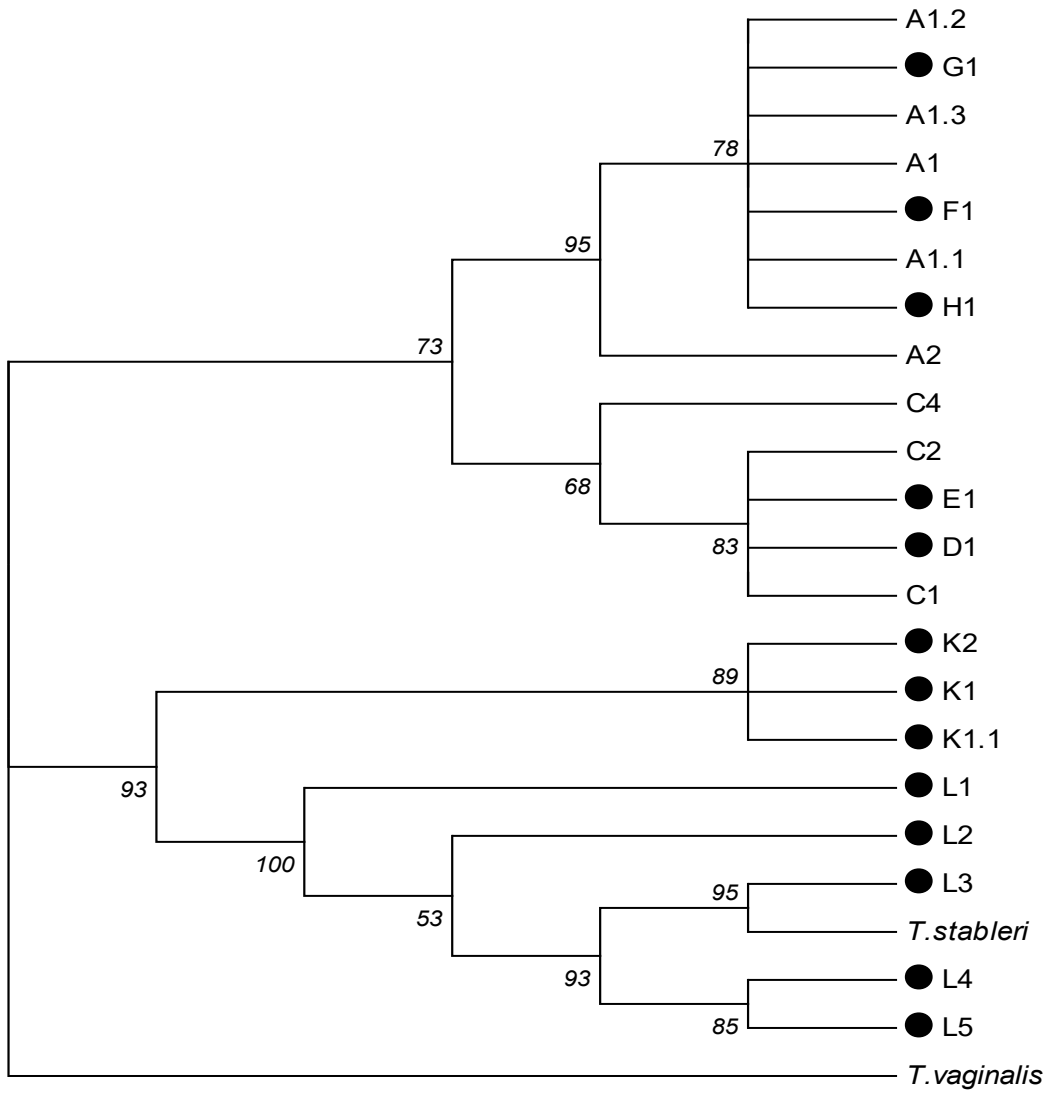

Seattle
WASHINGTON
Portland
OREGON
MONTANA
NORTH DAKOTA
SOUTH DAKOTA
IDAHO
WYOMING
NEBRASKA
NEVADA
UTAH
COLORADO
United States
Sacramento
San Francisco
San Jose
CALIFORNIA
Las Vegas
Los Angeles
San Diego
ARIZONA
Phoenix
Tucson
NEW MEXICO
El Paso
BAJA CALIFORNIA
SONORA
CHIHUAHUA
COAHUILA
NUEVO LEÓN
Gulf of California
Kansas City
KANSAS
OKLAHOMA
TEXAS
Dallas
Austin
San Antonio
Houston
MINNESOTA
Minneapolis
WISCONSIN
IOWA
Chicago
ILLINOIS
MISSOURI
ARKANSAS
MISSISSIPPI
LOUISIANA
New Orleans
MICHIGAN
Detroit
INDIANA
Indianapolis
OHIO
KENTUCKY
TENNESSEE
ALABAMA
GEORGIA
FLORIDA
Atlanta
Jacksonville
Orlando
Tampa
Charlotte
NORTH CAROLINA
SOUTH CAROLINA
Toronto
Ottawa
Montréal
Québec City
NEW YORK
PENNSYLVANIA
New York
Philadelphia
NJ
MARYLAND
DE
Washington
WEST VIRGINIA
VIRGINIA
VERMONT
MASSACHUSETTS
CT
RI

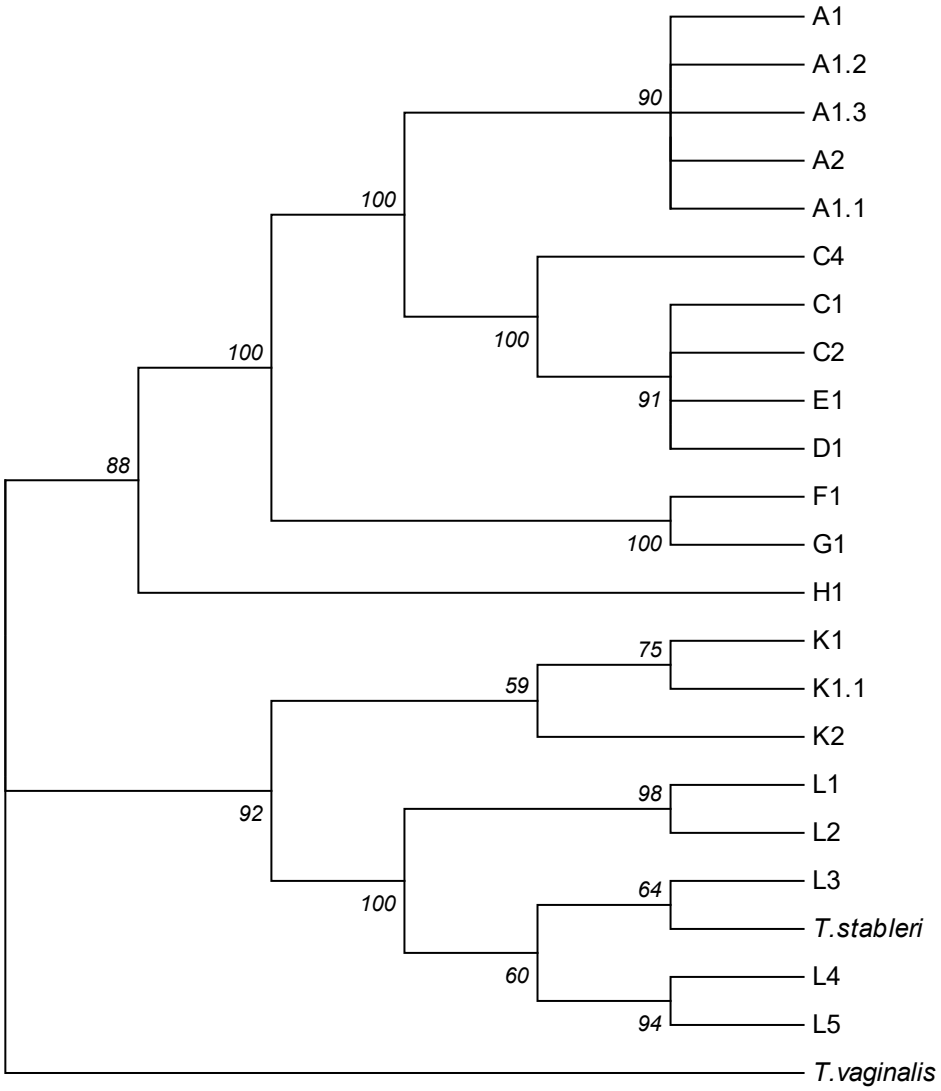
Figure 2


63
20
91
40
98
85
47
51
44
92
A
B
D
C
E
F
G
H
I
L
J
K (=T. stableri)
T.vaginalis

A

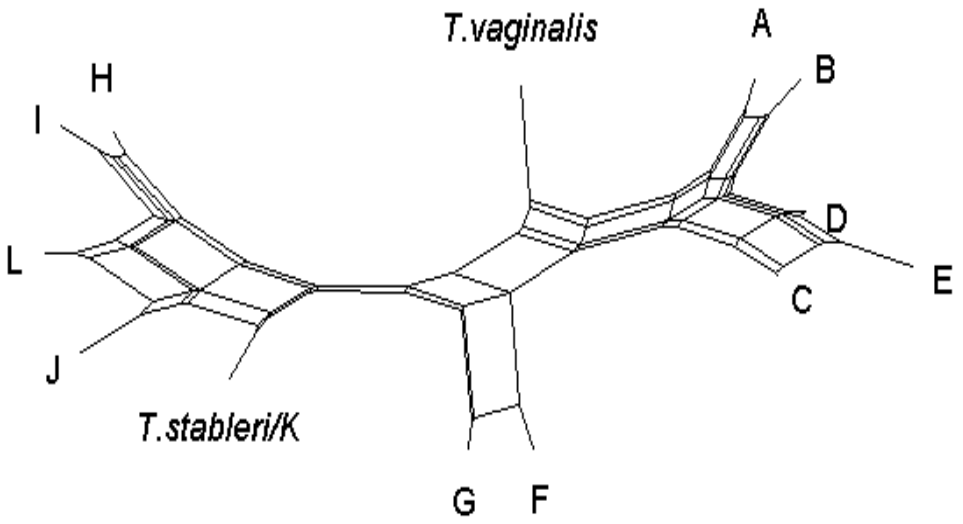

A1.2
G1
A1.3
78
A1
F1
95
A1.1
H1
73
A2
C4
C2
68
E1
83
D1
C1
K2
89
K1
K1.1
93
L1
L2
100
L3
95
53
T.stableri
93
L4
85
L5
T.vaginalis

B

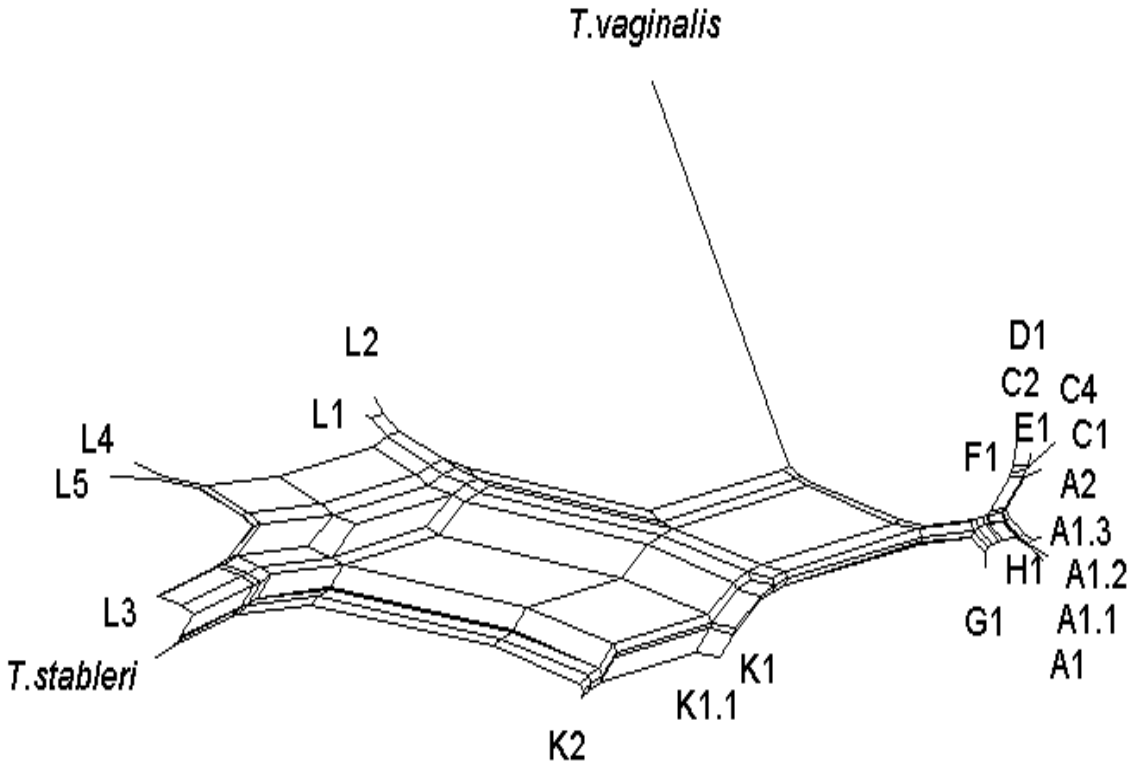

A1
A1.2
A1.3
A2
A1.1
C4
C1
C2
E1
D1
F1
G1
H1
K1
K1.1
K2
L1
L2
L3
T.stableri
L4
L5
T.vaginalis
90
100
100
100
91
88
100
75
59
98
92
64
100
60
94

C

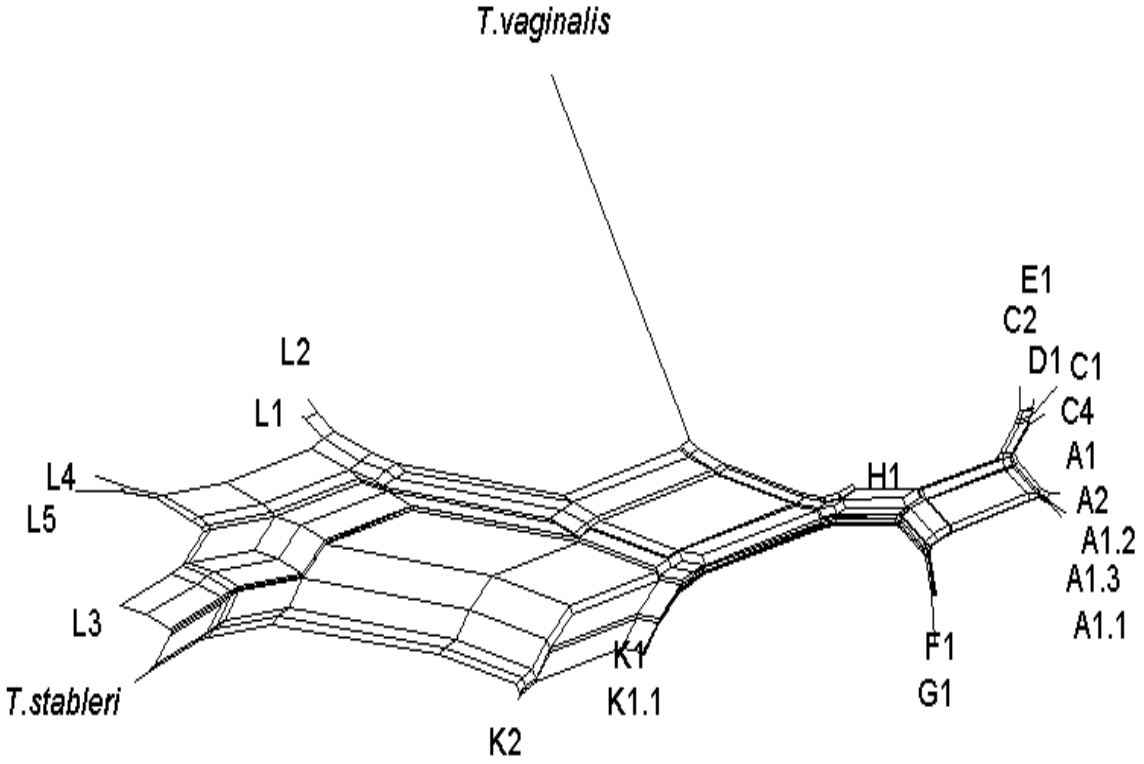
Figure 3


T.vaginalis
A
B
H
I
D
L
E
C
J
T.stableri/K
G
F

A


T.vaginalis
L2
D1
C2
C4
L1
E1
C1
L4
F1
L5
A2
A1.3
H1
A1.2
L3
G1
A1.1
T.stableri
K1
A1
K1.1
K2

B


T.vaginalis
E1
C2
D1
C1
C4
A1
A2
A1.2
A1.3
A1.1
L2
L1
L4
L5
L3
T.stableri
H1
F1
G1
K1
K1.1
K2

C