International Journal for Parasitology 50 (2020) 227–233

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

Whole genome sequencing of a feline strain of *Tritrichomonas foetus* reveals massive genetic differences to bovine and porcine isolates

Joanna Dąbrowska [a], Irene Keller [b], Jacek Karamon [a], Maciej Kochanowski [a], Bruno Gottstein [c], Tomasz Cencek [a], Caroline F. Frey [d], Norbert Müller [d,*]

[a] Department of Parasitology and Invasive Diseases, National Veterinary Research Institute, Partyzantów Avenue 57, 24-100 Puławy, Poland
[b] Department for BioMedical Research and Swiss Institute of Bioinformatics, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland
[c] Institute for Infectious Diseases, University of Bern, Friedbühlstrasse 51, 3010 Bern, Switzerland
[d] Institute of Parasitology, Vetsuisse Faculty, University of Bern, Länggassstrasse 122, 3012 Bern, Switzerland

ARTICLE INFO

Article history:
Received 16 August 2019
Received in revised form 17 December 2019
Accepted 18 December 2019
Available online 26 February 2020

Keywords:
Tritrichomonas foetus
Whole genome sequencing
Indel
Genotyping
Cat
Cattle
Pig

ABSTRACT

Tritrichomonas foetus is a protozoan parasite that colonizes the reproductive tract of cattle as well as the gastrointestinal tract of cats. Bovine tritrichomonosis is a sexually transmitted disease whereas feline tritrichomonosis is thought to be transmitted by the fecal-oral route. Furthermore, *T. foetus* is known as an essentially apathogenic commensal located in the nasal cavity of pigs. Transmission of *T. foetus* between the different hosts has to be considered a realistic scenario that may have important implications for the epidemiology of infections and disease. In our study, we generated whole genome sequencing (WGS) data from bovine, feline and porcine *T. foetus* strains to investigate the genetic (dis)similarities among these diverse strains. As a reference, we used a previously released draft assembly from a bovine *T. foetus* strain K isolated from an infected bull in Brazil. In particular, we identified single nucleotide polymorphisms (SNPs) and the insertion-deletion (indel) variations within the genomes of the different strains. Interestingly, only a low degree of polymorphism (68 SNPs and indels) was found between the bovine and the porcine strains in terms of variants with a predicted impact of moderate or high and where one species is homozygous for one allele and the other homozygous for the other allele. Conversely, however, a 964 times higher number of such differences was detected by comparing the feline with either the bovine (65,569) or the porcine (65,615) strain. These data clearly indicated a close phylogenetic relationship between bovine and porcine *T. foetus* but a remarkable genetic distinctness of these two strains from the feline strain. The latter observation was confirmed by PCR-based sequencing of 20 in silico-selected indel markers and five in silico-selected SNP markers that uniformly demonstrated a relatively distant phylogenetic relationship of three independent feline *T. foetus* isolates in comparison to the bovine and porcine strains investigated. In summary, our comparative genome sequencing approach provided further insights into the genetic diversity of *T. foetus* in relation to the different host origins of the parasite. Furthermore, our study identified a large number of SNP- and indel-containing sequences that may be useful molecular markers for future epidemiological studies aimed at the elucidation of the transmission patterns of *T. foetus* within different host species.

1. Introduction

Tritrichomonas foetus is the causative agent of bovine tritrichomonosis. This venereally transmitted protozoal parasite colonizes the mucosal surfaces of the urogenital tract of both male and female animals. In cows, infection with *T. foetus* causes apoptosis of vaginal epithelial cells and leads to vaginitis, cervicitis, endometritis and pyometra often associated with infertility or abortion in the early stage of pregnancy (BonDurant, 1997). Conversely, bulls infected with *T. foetus* remain asymptomatic. However, these animals are mostly persistently infected and represent the main reservoir for the parasite (Parker et al., 2001).

Bovine tritrichomonosis is on the list of infectious diseases notified by the World Organization of Animal Health (Trichomonosis, 2013; Health, 2015), and in many countries the disease is included in the list of infectious diseases subject to notification. In beef and

* Corresponding author.
E-mail address: norbert.mueller@vetsuisse.unibe.ch (N. Müller).

https://doi.org/10.1016/j.ijpara.2019.12.007
0020-7519/

dairy cattle farming, tritrichomonosis can cause a significant economic loss due to infertility of the animals, abortion, or culling of parasite carriers, respectively (Rae et al., 1999, 2004; Rodning et al., 2008). This is especially the case in South Africa, and countries from North, Central and South America where natural mating is still a predominant and common strategy in the breeding of cattle (Mukhufhi et al., 2003; Mendoza-Ibarra et al., 2012; Yao, 2013. In many parts of the world, however, bovine tritrichomonosis could be eliminated because artificial insemination with confirmed *T. foetus*-free semen had essentially replaced natural mating, thus basically avoiding transmission of the parasite during animal reproduction (Campero and Gottstein, 2007). Conversely, in countries where bovine cases of tritrichomonosis still occasionally arise such as Argentina, the Czech Republic and Australia, strict regulations and examinations to eradicate, or at least prevent spreading of, the disease are very important.

Several years ago, *T. foetus* infections in cats were identified as an issue of high clinical relevance, occurring in many geographic regions in the world with a prevalence ranging from 2 to 59% (Gookin et al., 2017). Feline tritrichomonosis is an intestinal disease of domestic cats, preferentially affecting those animals with an inbreeding background. In cats, *T. foetus* colonizes mucosal surfaces from the ileum, cecum, and colon, and causes neurophilic colitis, lymphocytic inflammation and chronic large bowel diarrhea. The parasite is supposed to be transmitted via the fecal-oral route and the risk of infection seems to be increased in an environment with a dense cat population as e.g. it is found in catteries (Gould et al., 2017).

T. foetus was also found in pigs and was identified as a commensal parsite both in the nasal cavity and the intestine of infected individuals. Porcine *T. foetus* was formerly termed as *Tritichomonas suis* and nowadays *T. foetus* and *T. suis* are generally considered as one species (e.g. reviewed in Frey and Müller, 2012). Since *T. foetus* is non-pathogenic in the porcine host, very little is known from the literature about the prevalence of this infection. Experimental cross-infections in different species showed that porcine *T. foetus* isolates are able to cause disease in cattle (Cobo et al., 2001). Similarly, the feline and bovine isolates were shown to cause at least mild symptoms in their reciprocal hosts (Stockdale et al., 2007, 2008).

Recent genotyping approaches involving PCR-based sequencing of various genetic markers revealed only a moderate genetic distinctness between the feline, bovine, and porcine genotypes of *T. foetus,* suggesting that they may represent the same species. In fact, multilocus genotyping of an axenic pig isolate of *T. foetus* confirmed the quasi-identity of this isolate to the bovine genotype. Furthermore, *T. foetus* from bovine and feline origin exhibited only a limited degree of conserved nucleotide polymorphism by comparative analyses of different genetic markers including cysteine protease 8 (CP8), internal transcribed spacer 2 (ITS 2) from rDNA region and elongation factor 1-alpha (Reinmann et al., 2012; Šlapeta et al., 2012; Sun et al., 2012). However, a very recent multilocus analysis identified more striking genetic differences between *T. foetus* from cats and cattle (Pedraza-Díaz et al., 2019). Nevertheless, spill-over of these apparently closely related genotypes from one to another host species have to considered a realistic scenario, although evidence for interspecies transmission of the parasite in natural environments is still lacking.

Recently, a draft whole genome sequence (WGS) GenBank™ accession number ASM183968v1) from a reference bovine strain of *T. foetus* (*T. foetus*, strain K, isolated in Brazil from an infected bull) was published by Benchimol et al. (2017). In the present study, we generated WGS data from bovine, porcine, and feline strains of *T. foetus* as a basis for subsequent comparative sequence analyses to assess - on an extended genomic level – the phylogenetic relationship between these strains.

2. Material and methods

2.1. Strains of T. foetus

In the study, bovine *T. foetus* strain ATCC 30924 (named bovine TF1), porcine strain ATCC 30167 (porcine TF3) and feline *T. foetus* strain IPA-Bern 2145 (feline TF2) isolated from a diarrheic cat at the Institute of Parasitology in Bern, Switzerland (Šlapeta et al., 2012) were used for WGS. For comparative indel and single nucleotide polymorphism (SNP) PCR marker evaluation (see Section 2.5), additional feline *T. foetus* strains DP-Puławy-16 and DP-Puławy-72 isolated from diarrheic cats at the Department of Parasitology, National Veterinary Research Institute (NVRI) in Puławy, Poland were tested. All strains originated from cats with tritrichomonosis were confirmed by real time PCR (Frey et al., 2017) and subsequent Sanger sequencing of the amplification products.

2.2. DNA extraction

In vitro cultivation of *T. foetus* trophozoites and extraction of genomic DNA from these cultures was done according to Frey et al. (2017). DNA concentration and purity of DNA samples was measured by the Quantifluor®dsDNA system (Promega).

2.3. Preparation of genomic DNA libraries and WGS

For WGS, genomic DNA libraries of TF1, TF2, and TF3 were prepared using Illumina TruSeq Nano (porcine and feline *T. foetus* strain TF2 and TF3) or Illumina TruSeq PCR-free (bovine *T. foetus* strain TF1) protocols according to the manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq 3000 at the NGS platform of the University of Bern to produce 2x150 bp reads. Raw sequencing data were deposited in the European Nucleotide Archive (ENA) under project accession number PRJEB31517 (https://www.ebi.ac.uk/ena).

2.4. Genome mapping and variant analyses

Reads were mapped to the *T. foetus* draft assembly ASM183968v1 (Benchimol et al., 2017) using Bowtie2 v. 2.3.4.1 in end-to-end mode. Picard-tools v. 2.2.1 was used to mark duplicates. Mapping rates were assessed with samtools flagstat v. 1.4 and average read depths with DepthOfCoverage from the Genome Analysis Toolkit (GATK) v. 3.7.

We used the GATK HaplotypeCaller to call variants (SNPs and small indels) separately for each sample and GATK GenotypeGVCF for joint calling across the three samples. Variants were quality filtered by applying cut-offs on various quality measures using GATK VariantFiltration and annotated using SnpEff v. 4.3 T with a database built from the annotation file for the *T. foetus* assembly downloaded from NCBI. The entire analysis workflow is illustrated in Supplementary Fig. S1. In order to retain only the most serious predicted effect for each variant, annotations were simplified with GATK VariantAnnotator.

2.5. Identification and design of indel and SNP PCR markers

Based on the WGS data, we developed a PCR assay to discriminate the feline *T. foetus* strain from the other two isolates. For this purpose, candidate indel- (Table 1, Supplementary Fig. S2) and SNP- (Table 2; Supplementary Fig. S2) containing regions were inspected using Integrative Genomics Viewer (IGV) software

Table 1
Chromosomal localization of selected insertions/deletions (indels) within the draft whole genome sequence of feline *Tritrichomonas foetus* strain TF2 and annotation of affected proteins as compared to the reference genome from bovine *T. foetus* K strain (GenBank™ accession number ASM183968v1).

No	Chromosome	Position	Gene symbol	Protein annotation
1	MLAK01000606.1	56,272	TRFO 20016	hypothetical protein TRFO 20016 [OHT10611]
2	MLAK01000046.1	63,119	TRFO 12728	DnaJ domain containing protein TRFO 12728 [OHT17029.1]
3	MLAK01000217.1	148,844	TRFO 02862	hypothetical protein TRFO 02862 [OHT15531.1]
4	MLAK01000272.1	54,769	TRFO 14444	hypothetical protein TRFO 14444 [OHT15094.1]
5	MLAK01000303.1	98,157	TRFO 14711	hypothetical protein TRFO 14711 [OHT14913.1]
6	MLAK01000013.1	12,115	TRFO 41109	iron hydrogenase TRFO 41109 [OHT17348.1]
7	MLAK01000605.1	63,177	TRFO 04144	hypothetical protein TRFO 4144 [OHT10644.1]
8	MLAK01000605.1	3557	TRFO 04140	adaptin N terminal region family protein [OHT10642.1]
9	MLAK01000605.1	133,776	TRFO 04181	hypothetical protein TRFO_04181 [OHT10674.1]
10	MLAK01000545.1	53,455	TRFO 16686	hypothelical protein TRFO 16686 [OHT13240.1]
11	MLAK01000111.1	39,983	TRFO 02644	hypothetical protein TRFO 02644 [OHT16405.1]
12	MLAK01000111.1	92,196	TRFO 02675	CAMK family protein kinase [OHT16420.1]
13	MLAK01000111.1	110,478	TRFO 02684	hypothetical protein TRFO_02684 [OHT16428.1]
14	MLAK01000560.1	165,801	TRFO 03634	small GTP-binding protein TRFO 03634 [OHT12604.1]
15	MLAK01000111.1	156,322	TRFO 02707	hypothetical protein TRFO_02707 [OHT16445.1]
16	MLAK01000014.1	9390	TRFO 12428	hypothetical protein TRFO_12428 [OHT17303.1]
17	MLAK01000014.1	19,536	TRFO 12437	hypothetical protein TRFO_12437 [OHT17303.1]
18	MLAK01000001.1	63,599	TRFO_00893	hypothetical protein TRFO_00893 [OHT17625.1]
19	MLAK01000002.1	25,060	TRFO_02470	hypothetical protein TRFO_02470 [OHT17460.1]
20	MLAK01000112.1	116,771	TRFO_13254	hypothetical protein TRFO_13254 [OHT16392.1]

Table 2
Chromosomal localization of selected single nucleotide polymorphism (SNPs) within the draft whole genome sequence of feline *Tritrichomonas foetus* strain TF2 and annotation of affected proteins as compared to the reference genome from bovine *T. foetus* K strain (GenBank™ accession number ASM183968v1).

No	Chromosome	Position	Gene symbol	Protein annotation
1	MLAK01000001.1	2858	TRFO_00788	dynein heavy chain family protein TRFO_00788 [OHT17559.1]
2	MLAK01000325.1	88,026	TRFO_02993	hypothetical protein TRFO_02993 TRFO_02993 [OHT14734.1]
3	MLAK01000335.1	68,707	TRFO_14933	hypothetical protein TRFO_14933 [OHT14635.1]
4	MLAK01000217.1	45,925	TRFO_02793	hypothetical protein TRFO_02793 [OHT15488.1]
5	MLAK01000229.1	92,678	TRFO_14151	hypothetical protein TRFO_14151 [OHT15393.1]

Table 3
Sizes of selected insertion/deletion (indel) regions and parameters (primers sequences, primer annealing temperatures, and expected amplification product sizes) used for PCR-based validation of indel variations between the draft whole genome sequence of feline *Tritrichomonas foetus* strain TF2 and the reference genome from bovine *T. foetus* K strain (GenBank™ accession number ASM183968v1).

No	Variation in TF2	Annealing temp.	Forward primer	Reverse primer	Product size for TF2 (in bp)
1	Deletion in cat (20 bp)	59 °C	CGCGAATCTTTCTTGTTTGCG	TGTGAAGCGTTTTAAAGTGAACTGA	283
2	Insertion in cat (10 bp)	57 °C	AATGCTGTGAACACTTTTCGTT	CAAAATTATTTTGCTGGTTAAGGTCA	144
3	Insertion in cat (13 bp)	57 °C	TTTCTGTTAGACGTATACCCGAA	TTGGAAAAAGAAACAGAACAAGAAAA	276
4	Insertion in cat (9 bp)	60 °C	GGCTCGAAAAAGGAAAGGGC	GAAGAGGAATGCAAAACGCGT	183
5	Deletion in cat (21 bp)	58 °C	TCACATTTCACGATCAATCATAGAAT	AGACTAGGTTCAAGAAAAGAAAAGTCA	141
6	Deletion in cat (24 bp)	57 °C	CCCGAAAACGCGAATTCCTC	GGATTTCCGATGGTGGTGGT	175
7	Deletion in cat (6 bp)	57 °C	TCGTTGCTGCTGATATTGCT	TGCGGTGTTGTACATGCTGA	241
8	Insertion in cat (9 bp)	60 °C	CCCTATCCGCCCTATCCGTA	GAGCAGGTTGAAGAGGCAGT	190
9	Deletion in cat (10 bp)	60 °C	AGTTCAACTGCCTGCTCTGG	ATATGAGCACCGGGTAGGGT	134
10	Deletion in cat (9 bp)	58 °C	AGCATCGTGAATAACAAAAACGT	TGATTCACCATTTTCACCAGCAA	180
11	Insertion in cat (9 bp)	59 °C	GCTAGCTAAATGGGGAAAACATGA	TCGATGTCCAATTCTGAGAGTGA	304
12	Insertion in cat (12 bp)	58 °C	TGAATCCGGAAAAGAACCACA	TGAATCTGTGATGTTCTCGCA	222
13	Deletion in cat (9 bp)	60 °C	AGCCTTTCGTCACTTTCCCA	AGCATGTTACTCCAAGATTATCGA	180
14	Deletion in cat (18 bp)	58 °C	AGGAGACTAACCCAATAATTGGCA	TGGTTTAATTATCCCCTTATTCTGC	224
15	Deletion in cat (12 bp)	57 °C	ACAAGTTGAACTTAGAAACAAGTTGA	ATGCATAAGGGTGGAAAGCA	196
16	Insertion in cat (18 bp)	57 °C	TCAAAACGTGTCCAGCATGC	ACGACTGTCCAAGTGGTTCC	521
17	Insertion in cat (7 bp)	60 °C	CCAACACTGGCAGCAATCAC	TCCCTTTGGGACCAGCTTGT	257
18	Insertion in cat (18 bp)	58 °C	ACTTCGACAGATACGAGCCA	TTTGATCTTTCCATTTTTGGCCT	307
19	Insertion in cat (17 bp)	57 °C	AGCAAATCCGCACAAAGATGA	TCCGCCAAATTTCGATGAAGC	523
20	Insertion in cat (15 bp)	57 °C	GCCAAAGTTGTCCAATGGGG	CCTGATGAATCTTTTCTAACCTACCA	356

(http://www.broadinstitute.org/igv) and these regions were selected for primer development (Tables 3 and 4).

PCR primers were designed with Primer3 in Geneious R7 software (Biomatters, http://www.geneious.com). Designed primer sets and parameters for PCR-based amplification of the 20 indel- and five SNP-containing regions selected in silico are shown in Tables 3 and 4, respectively.

2.6. PCR-based amplification of indel and SNP markers, and sequencing of amplification products

PCRs were performed in the Biometra Thermal Cycler (Biometra, Kent, UK) using the QIAGEN multiplex PCR kit (Qiagen, Hilden. Germany). PCR mixes contained 0.4 μM of each primer and 25–30 ng/μl of *T. foetus* DNA from cats, 48 ng/μl from cattle and 8.5 ng/μl from pigs.

Table 4
Sizes of selected single nucleotide polymorphism (SNP) regions and parameters (primers sequences, primer annealing temperatures, and expected amplification product sizes) used for PCR-based validation of SNP variations between the draft whole genome sequence of feline *Tritrichomonas foetus* strain TF2 and the reference genome from bovine *T. foetus* K strain (GenBank™ accession number ASM183968v1).

No	Number of SNP variations in TF2	Annealing temp.	Forward primer	Reverse primer	Product size for TF2 (in bp)
1	4	59 °C	CCTTGTCCTCTTCCTCGACC	TGTGTGGTGTCTTGTTACTGGT	426
2	4	58 °C	ACGACTTTTTGAATTGACTTCCGA	TGCATCCTGAATGAAGCATCA	434
3	4	59 °C	CCTGCGTGTTGGTTTTCCTC	CAAGAAACCAGCAGCAGCAG	379
4	7	58 °C	TCTCTGAACAGAAATATTGAATGCAA	GCTCGCCTGGATTTGAATCG	405
5	4	59 °C	GGAAATCCCCTTCCCCCATAT	GTTGAGATGAACTGGCCCCT	376

The PCRs were initiated by a 10 min incubation at 94 °C (activation of HotStar Taq DNA polymerase) and the reaction was continued in 35 cycles (denaturation (94 °C, 30 s), annealing (57 °C–60 °C, 30 s, see Tables 3 and 4), and extension (72 °C, 90 s)) followed by one cycle at 72 °C for 10 min. PCR amplification products were analyzed by agarose gel electrophoresis (2% gels stained with ethidium bromide (Sigma, St. Louis, USA)). Additionally, PCR amplification products were subjected to bi-directional Sanger sequencing using a commercial sequencing service (Genomed S.A. Company, Warsaw, Poland). For multiple sequence alignments the TCOFFEE::Regular programme (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi?s-tage1=1&daction=TCOFFEE::Regular) was used.

3. Results

3.1. Whole genome re-sequencing

Genomic DNA from *T. foetus* strains originating from cattle (bovine TF1), cats (feline TF2) and pigs (porcine TF3) were sequenced using the Illumina HiSeq 3000 sequencing platform and 131 million (M) (TF1), 162 M (TF2) and 484 M (TF3) raw pair-end reads were generated. Total alignment rates were 59.3% (bovine TF1), 81.5% (feline TF2), and 62.8% (porcine TF3), and the average read depths 227.7 (bovine TF1), 485.7 (feline TF2) and 971.3 (porcine TF3), respectively (Table 5).

3.2. Comparison of the bovine TF1, feline TF2, and porcine TF3 genomes to the reference genome

The numbers of variants (i.e. SNPs and indels) in the TF1, TF2, and TF3 genomes were compared with the bovine *T. foetus* reference genome GenBank™ accession number ASM183968v1) (Table 6). We distinguish between heterologous positions (i.e. one copy of the reference allele, one copy of the alternative allele) and homologous positions with two copies of the alternative allele (see also Discussion). A relatively low number of heterologous positions was determined in the case of bovine TF1 (3946) and porcine TF3 (4321), whereas feline TF2 demonstrated a much higher number of such positions (6183 counts). In the homologous situation, however, these differences were much more pronounced in that feline TF2 (338,632) exhibited >8 times more variants than bovine TF1 (39,784) and porcine TF3 (38,916), respectively. In summary, the data described above indicated that bovine TF1 and porcine TF3 were genetically closely related to the bovine *T. foetus* reference strain K, isolated in Brazil, whereas feline TF2 appeared to be genetically rather distinct.

Table 5
Summary statistics for whole genome sequencing (WGS) data of bovine TF1, feline TF2, and porcine TF3 *Tritrichomonas foetus* strains.

	Bovine TF1	Feline TF2	Porcine TF3
Total number of reads pairs	131,522,838	162,809,690	484,954,758
Total alignment rate to *T. foetus* draft assembly (%)	59.3	81.5	62.8
Mean read depth	227.7	485.7	971.3

Table 6
Number of variants (i.e. single nucleotide polymorphism (SNPs) or insertions/deletions (indels)) discriminating bovine TF1, feline TF2, and porcine TF3 *Tritrichomonas foetus* strains from the reference bovine *T. foetus* strain K (GenBank™-accession number ASM183968v1).

	Bovine TF1	Feline TF2	Porcine TF3
Heterologous (i.e. one copy of the reference allele, one copy of the alternative allele)	3946	6183	4321
Homologous for alternative (non-referenced) allele	39,784	338,632	38,916

3.3. Genome diversity between bovine, feline, and porcine T. foetus strains

Next, we identified the number of differences (SNPs and small indels) between isolates. As shown in Table 7, genetic differences were highly abundant between feline TF2 and bovine TF1 (338,803 variants) or porcine TF3 (338,916 variants), respectively. In contrast, genetic differences between bovine TF1 and porcine TF3 were extremely low (392 variants). The same patterns hold when we considered only protein-changing variants. Specifically, these included all variants where the effect on protein function was predicted to be moderate (e.g. in-frame variants leading to non-synonymous coding changes, codon deletion or insertion) or high (caused e.g. by out-of-frame variants leading to likely disruption of protein function). Here, 65,569 variants between feline TF2 and bovine TF1 (Table 7 and Supplementary Table S1) and 65,615 variants between feline TF2 and porcine TF3 (Table 7 and Supplementary Table S2) were identified. Conversely, only 68 TF1 versus TF3 variants with moderate to high impact on the corresponding protein functions were found in this comparative sequence analysis (Table 7 and Supplementary Table S3). In summary, these data again demonstrated the relatively close phylogenetic relationship between bovine TF1 and porcine TF3, whereas feline TF2 turned out to be genetically clearly distinct from these two strains.

Table 7
Genomic variation between bovine TF1, feline TF2, and porcine TF3 *Tritrichomonas foetus* strains.

	Feline TF2 versus bovine TF1	Feline TF2 versus porcine TF3	Bovine TF1 versus porcine TF3
Any type of variant[a]	338,803	338,916	392
Variant with predicted impact moderate[b] to high[c]	65,569	65,615	68

[a] Any variant within the entire genome.
[b] Moderate impact: a non-disruptive variant potentially changing protein effectiveness.
[c] High impact – a variant putatively disrupting protein function.

3.4. Comparative analysis of selected indel- and SNP-containing regions

Using the GATK HaplotypeCaller algorithm, 59,103 indel variations were identified between the reference genome from bovine *T. foetus* K strain and the genome of feline TF2. We considered such indels to be potential genetic markers that are suitable for a PCR genotyping approach to confirm the genetic polymorphism among bovine, feline, and porcine *T. foetus* strains TF1, TF2, and TF3 based on gel electrophesis, thus relying only on amplicon length without the need to sequence the amplicon. In order to prove this concept, 20 sequence stretches containing four to 24 bp indels were in silico-selected (see Table 3) as PCR targets for genotyping. Amplification of these indel regions and subsequent size determination by gel electrophoresis and confirmatory sequencing of the respective amplification products provided the following results: Swiss feline strain TF2 and two other feline isolates from Poland, DP-Puławy-16 and DP-Puławy-72 (not shown), exhibited indel DNA amplification products that were consistently distinct from those of bovine TF1 and porcine TF3, respectively. Conversely, indels among the three feline strains tested turned out to be overall identical. Here, an identical indel pattern was also observed for bovine TF1 and porcine TF3. In the case of indel number 5 (see Table 3), our approach was exemplified in that amplicons from feline TF2 versus bovine TF1/porcine TF3 demonstrated size differences consistent with the expected 21 bp indel (141 versus 162 bp) (Fig. 1). Clearly distinct amplicon patterns were also detected for other PCRs that were targeted to indels with insertions or deletions of ≥15 nucleotides (nts) (indels 1, 5, 6, 14, 16, 18, 19, 20, see Table 3) (not shown). For indels with insertions or deletions <15 nts (indels 2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 15, 17; see Table 3) discrimination relied on sequencing of the amplification products (not shown). Analogous results were achieved by PCR-based sequence comparison of five selected regions from bovine (TF1), feline (TF2), and porcine (TF3) *T. foetus* strains that contained 4–7 SNPs (23 SNPs in total) as identified by the GATK HaplotypeCaller algorithm (Table 4). In this analysis, amplification of SNP-containing regions and subsequent sequencing of the respective amplification products provided the following results: Swiss feline strain TF2 and two other feline isolates from Poland DP-Puławy-16 and DP- Puławy-72 (not shown) exhibited SNP-containing DNA amplification products that were consistently distinct from those of bovine TF1 and porcine TF3, respectively. Conversely, SNPs among the three feline strains tested turned out to be overall identical. Here, an identical SNP pattern was also observed for bovine TF1 and porcine TF3 (Supplementary Fig. S3).

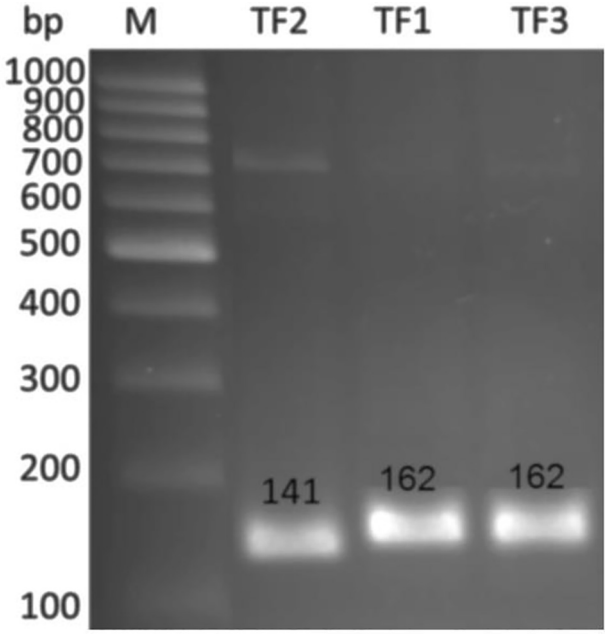

Fig. 1. As exemplified for insertion/deletion (indel) marker no. 5 (see Table 3), PCR amplification products of the respective genomic indel regions from feline TF2, bovine TF1, and porcine TF3 *Tritrichomonas foetus* strains were separated by 2% agarose gel electrophoresis. Size markers (M) (on the left) and expected fragment lengths of amplification products (above the DNA bands) are given in base pairs (bp). The down-shift of the feline TF2 versus the bovine TF1 and porcine TF3 amplification products is consistent with the expected 21 bp deletion in the indel region of TF2.

In summary, PCR-based sequencing analysis of the 20 indel and 23 SNPs distributed over five independent DNA regions markers (see above) confirmed the close genetic relationship among all feline *T. foetus* strains on the one hand and among the bovine and porcine *T. foetus* strains on the other hand. In contrast, a clear genetic distinctness between the feline group of *T. foetus* strains on one side and bovine TF1 and porcine TF3 on the other side was observed. Furthermore, since all in silico-selected indels and SNPs turned out to be "true" variants in our PCR-based sequencing analyses, we can conclude that application of the GATK HaplotypeCaller algorithm alone was a successful and fully trustworthy variant caller method to demonstrate the genetic differences between the feline, bovine, and porcine strain of *T. foetus* on the WGS level.

4. Discussion

Recent multilocus genotyping approaches revealed only a minor genetic distinctness between feline, bovine, and porcine strains of *T. foetus,* suggesting that they may belong to the same species. In particular, such studies comparing bovine and porcine *T. foetus* (Tachezy et al., 2002; Šlapeta et al., 2012) as well as bovine and feline *T. foetus* isolates (Reinmann et al., 2012; Šlapeta et al., 2012; Sun et al., 2012) yielded only minor genetic variation among the different isolates investigated. However, although these studies included several methods such as restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), and/or PCR-based sequencing analysis of multiple genetic markers for genotyping, and seemed to be highly conclusive, we were still wondering whether they could sufficiently resolve the phylogenetic status of *T. foetus* from cattle, pigs, and cats.

In order to generate additional information about the phylogeny of *T. foetus*, we investigated on the entire genome level the genetic diversity of *T. foetus* in relation to the different host origins of the parasite. We chose this sophisticated approach for three main reasons: (i) our WGS approach was facilitated by the availability of a draft genome sequence from a bovine *T. foetus* strain K (Benchimol et al., 2017) serving as a scaffold for whole genome re-sequencing of our bovine, feline and porcine strain of *T. foetus;* (ii) WGS represents the "gold standard" method for determination of the phylogenetic relationship between different organisms; (iii) by generating respective WGS data we could provide an ideal basis for the identification of novel genetic markers such as indels suitable for phylogenetic investigation of *T. foetus* of different host origins.

Our WGS-based SNP and indel analyses confirmed previous results suggesting a very close phylogenetic relationship between bovine and porcine *T. foetus*. Accordingly, our findings further supported the assumption that *T. foetus* originating from these two hosts are co-specific (Tachezy et al., 2002). Thus - at least from the genetic point of view – they may have to be considered as potentially infectious, and perhaps also pathogenic, in their reciprocal hosts. Conversely however, it also became evident that in fact

abundant genetic differences do exist between feline and bovine strains as well as between feline and porcine *T. foetus* strains. These genetic differences included a large number of variants that have a predicted impact on the coding sequences of the different strains. This striking genetic distinctness of the feline strains was rather surprising, referring to previous genotypic evaluations that suggested a low degree of genetic polymorphism among all *T. foetus* strains irrespective of their host origin (Reinmann et al., 2012; Šlapeta et al., 2012; Sun et al., 2012). However, a closer look at proteomic and transcriptomic data led to a different conclusion. For example, in two-dimensional gel-electrophoretic zymography using protease-specific fluorogenic substrates, marked differences in cysteine protease (CP) expression profiles between the two genotypes were observed (Stroud et al., 2017). Furthermore, transcriptomic data e.g. considering a comparative sequence alignment among the transcripts from a family of protease inhibitors, exhibited a genetic variation of 1–2% between the bovine and feline strains tested (Morin-Adeline et al., 2015) indicating a relatively high genetic variability between these two genotypes. This was confirmed in a very recent multilocus analysis that identified more striking genetic differences between *T. foetus* from cats and cattle (Pedraza-Díaz et al., 2019).

By performing WGS with bovine, porcine, and feline *T. foetus* strains, we also intended to provide a basis for the identification of new genetic markers that may allow a refined investigation of the polymorphism of *T. foetus*. In this respect, the 20 indels selected for detailed analysis in the present study (see Table 3) turned out to be highly suitable as genetic markers to unambiguously distinguish feline from bovine and porcine strains. Accordingly, genotyping using PCRs targeting these indels can be considered an ideal molecular epidemiological concept to explore the spill-over theory, taking into consideration a transmission of *T. foetus* from cats to cattle or pigs (or vice versa), respectively. Importantly, the use of relatively large indels (≥15 nt such as indels No. 1, 3, 5, 6, 14, 16, 19, and 20; see Table 3) for the realization of this concept would allow large-scale epidemiological investigations by simple agarose gel electrophoresis and assessment of the amplicon sizes (see Fig. 1) instead of elaborate and expensive sequencing of the PCR amplification products. Conversely, our study was not focused on the identification of genetic markers that could monitor subgenotypes within groups of *T. foetus* strains selected according to their host origin. For this purpose, adequate genetic markers may be identified among those few SNPs and indels discriminating the bovine from the porcine *T. foetus* strains (see Supplementary Table S3).

Compared with other parasitic protists, *T. foetus* (similar to other trichomonad species) habors an extremely large genome (ca. 68.5 Mb) that is characterized by superabundant repeated sequences and a remarkably high number of gene duplications (Zubácová et al., 2008; Oyhenart and Breccia, 2014; Benchimol et al., 2017). Due to this complex situation, the elucidation of the *T. foetus* genome structure remains a major scientific issue in the field of tritrichomonosis. In this context, for example the unsolved discrepancy in cytogenetic karyotype definition (diploid versus haploid genome) (Xu et al., 1998; Zubácová et al., 2008) causes difficulties in discriminating between heterozygous, orthologous or paralogous forms of gene duplication within the genome. In the present study, large numbers of heterologous (i.e. two distinct alleles of a given gene) and homologous (i.e. two identical alleles of a given gene) gene duplications were identified in the bovine, feline and porcine strains of *T. foetus* (see Table 6). Unfortunately, however, our draft WGS sequence data were also not suited to characterize in detail the architecture of the *T. foetus* genome and to resolve the ploidy pattern of the parasite. Accordingly, our study did not allow any conclusions to be drawn with respect to the heterozygosity, orthology or paralogy status of the gene duplications identified in the different *T. foetus* genomes analyzed.

An important purpose of our study was to generate a large amount of *T. foetus* genomic sequence information to provide a solid basis for further genotyping approaches in the field of tritrichomonosis. Although our study clearly indicated remarkable genetic differences between feline *T. foetus* strains and bovine or porcine strains, respectively, WGS data per se are insufficient to justify alterations in the taxonomy and nomenclature within the *Tritrichomonas* spp. This is particularly not the case, because at least to our knowledge, unambiguous criteria as basis for the definition of a new species within the genera of trichomonads and related organisms are not yet available. Apart from that, in our study, genetic variation was not demonstrated for a representative set of relevant *T. foetus* strains.

In summary, our WSG approach provided important insights into the genetic diversity of *T. foetus* in relation to the different host origins of the parasite. Furthermore, a large number of SNPs and indels were identified that may serve as molecular markers for future epidemiological studies aiming to elucidate the transmission patterns of *T. foetus* within different host species. Finally, our WGS data may also serve as a valuable basis to apply the entire 'omic' technology to elucidate those cellular functions that contribute to both the apparent host specificities and the pathogenicity patterns of the bovine, porcine, and feline *T. foetus* strains.

Acknowledgements

The authors wish to thank Muriel Fragnière and Sabrina Schenk for expert technical assistance, Walter Basso for creating the graphical abstract (copyright assigned to the authors), and Andrew Hemphill for carefully reading the manuscript. The Next Generation Sequencing Platform and the Interfaculty Bioinformatics Unit of the University of Bern, Switzerland are acknowledged for performing the whole genome re-sequencing experiments and providing high performance computing infrastructure. This study was supported by a Swiss Government Excellence Scholarship for Foreign Students. This paper is part of Joanna Dąbrowska's Ph.D. dissertation (National Veterinary Research Institute, Puławy, Poland).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2019.12.007.

References

Benchimol, M., de Almeida, L.G., Vasconcelos, A.T., de Andrade Rosa, I., Bogo, M.R., Kist, L.W., de Souza, W., 2017. Draft genome sequence of *Tritrichomonas foetus* strain K. Genome Announc. 5, e00195–00117.

BonDurant, R.H., 1997. Pathogenesis, diagnosis, and management of trichomoniasis in cattle. Vet. Clin. Food Anim. 13, 345–361.

Campero, C.M., Gottstein, B., 2007. Tritrichomonosis. In: Ortega-Mora, L.M., Gottstein, B., Conraths, F.J., Buxton, S. (Eds.), Protozoal Abortion in Farm Ruminants: Guidelines for Diagnosis and Control. Athenaeum Press, Gateshead, UK, pp. 290–301.

Cobo, E.R., Cano, D., Campero, C.M., 2001. Experimental infection with *Tritrichomonas suis* in heifers. 2001. Vet. Parasitol. 99, 73–78.

Frey, C.F., Müller, N., 2012. *Tritrichomonas* - systematics of an enigmatic genus. Mol. Cell. Probes. 26, 132–136.

Frey, C.F., Müller, N., Stauber, N., Marreros, N., Hofmann, L., Hentrich, B., Hirsbrunner, G., 2017. *Simplicimonas*-like DNA in vaginal swabs of cows and heifers cross-reacting in the real-time PCR for *T. foetus*. Vet. Parasitol. 237, 30–36.

Gookin, J.L., Hanrahan, K., Levy, M.G., 2017. The conundrum of feline trichomonosis. J. Feline Med. Surg. 19, 261–274.

Gould, E.N., Giannone, R., Kania, S.A., Tolbert, M.K., 2017. Cysteine protease 30 (CP30) contributes to adhesion and cytopathogenicity in feline *Tritrichomonas foetus*. Vet. Parasitol. 244, 114–122.

Health, W.O.F.A., 2015. OIE-Listed Diseases, Infections and Infestations in Force in 2017. OIE France.
Mendoza-Ibarra, J.A., Pedraza-Díaz, S., García-Peña, F.J., Rojo-Montejo, S., Ruiz-Santa-Quiteria, J.A., San Miguel-Ibáñez, E., Navarro-Lozano, V., Ortega-Mora, L. M., Osoro, K., Collantes-Fernandez, E., 2012. High prevalence of *Tritrichomonas foetus* infection in Asturiana de la Montaña beef cattle kept in extensive conditions in Northern Spain. Vet. J. 193, 146–151.
Morin-Adeline, V., Mueller, K., Conesa, A., Slapeta, J., 2015. Comparative RNA-seq analysis of the *Tritrichomonas foetus* PIG30/1 isolate from pigs reveals close association with *Tritrichomonas foetus* BP-4 isolate 'bovine genotype'. Vet. Parasitol. 212, 111–117.
Mukhufhi, N., Irons, P.C., Michel, A., Peta, F., 2003. Evaluation of a PCR test for the diagnosis of *Tritrichomonas foetus* infection in bulls: effects of sample collection method, storage and transport medium on the test. Theriogenology 60, 1269–1278.
Oyhenart, J., Breccia, J.D., 2014. Evidence for repeated gene duplications in *Tritrichomonas foetus* supported by EST analysis and comparison with the *Trichomonas vaginalis* genome. Vet. Parasitol. 206, 267–276.
Parker, S., Lun, Z.R., Gajadhar, A., 2001. Application of a PCR assay to enhance the detection and identification of *Tritrichomonas foetus* in cultured preputial samples. J. Vet. Diagn. Invest. 13, 508–513.
Pedraza-Díaz, S., Arranz-Solís, D., Gómez-Couso, H., Fuschs, L., Fort, M., Rengifo-Herrera, C., Navarro-Lozano, V., Ortega-Mora, L.M., Collantes-Fernández, E., 2019. Multilocus analysis reveals further genetic differences between *Tritrichomonas foetus* from cats and cattle. Vet. Parasitol. 276, 108965.
Rae, D., Chenoweth, P., Genho, P., McIntosh, A., Crosby, C., Moore, S., 1999. Prevalence of *Tritrichomonas foetus* in a bull population and effect on production in a large cow-calf enterprise. J. Am. Vet. Med. Assoc. 214, 1051–1055.
Rae, D.O., Crews, J.E., Greiner, E.C., Donovan, G.A., 2004. Epidemiology of *Tritrichomonas foetus* in beef bull populations in Florida. Theriogenology 61, 605–618.
Reinmann, K., Müller, N., Kuhnert, P., Campero, C.M., Leitsch, D., Hess, M., Henning, K., Fort, M., Müller, J., Gottstein, B., Frey, C.F., 2012. *Tritrichomonas foetus* isolates from cats and cattle show minor genetic differences in unrelated loci ITS-2 and EF-1alpha. Vet. Parasitol. 185, 138–144.
Rodning, S.P., Wolfe, D.F., Carson, R.L., Wright, J.C., Stockdale, H.D., Pacoli, M.E., Busby, H.C., Rowe, S.E., 2008. Prevalence of *Tritrichomonas foetus* in several subpopulations of Alabama beef bulls. Theriogenology 69, 212–217.
Šlapeta, J., Müller, N., Stack, C.M., Walker, G., Lew-Tabor, A., Tachezy, J., Frey, C.F., 2012. Comparative analysis of *Tritrichomonas foetus* (Riedmüller, 1928) cat genotype, *T. foetus* (Riedmüller, 1928) cattle genotype and *Tritrichomonas suis* (Davaine, 1875) at 10 DNA loci. Intl. J. Parasitol. 42, 1143–1149.
Stockdale, H.D., Dillon, A.R., Newton, J.C., Bird, R.C., Bondurant, R.H., Deinnocentes, P., Barney, S., Bulter, J., Land, T., Spencer, J.A., Lindsay, D.S., Blagburn, B.L., 2008. Experimental infection of cats (*Felis catus*) with *Tritrichomonas foetus* isolated from cattle. Vet. Parasitol. 154, 156–161.
Stockdale, H., Rodning, S., Givens, M., Carpenter, D., Lenz, S., Spencer, J., Dykstra, C., Lindsay, D., Blagburn, B., 2007. Experimental infection of cattle with a feline isolate of *Tritrichomonas foetus*. J. Parasitol. 93, 1429–1434.
Stroud, L.J., Slapeta, J., Padula, M.P., Druery, D., Tsiotsioras, G., Coorssen, J.R., Stack, C. M., 2017. Comparative proteomic analysis of two pathogenic *Tritrichomonas foetus* genotypes: there is more to the proteome than meets the eye. Int. J. Parasitol. 47, 203–213.
Sun, Z., Stack, C., Šlapeta, J., 2012. Sequence differences in the diagnostic region of the cysteine protease 8 gene of *Tritrichomonas foetus* parasites of cats and cattle. Vet. Parasitol. 186, 445–449.
Tachezy, J., Tachezy, R., Hampl, V., Sedinova, M., Vanacova, S., Vrlik, M., Van Ranst, M., Flegr, J., Kulda, J., 2002. Cattle pathogen *Tritrichomonas foetus* (Riedmuller, 1928) and pig commensal *Tritrichomonas suis* (Gruby & Delafond, 1843) belong to the same species. J. Eukaryot. Microbiol. 49, 154–163.
Trichomonosis, O., 2013. OIE Terrestrial Animal Health Code 2. World Organisation for animal Health, Paris, France, p. 574.
Xu, W.D., Lun, Z.R., Gajadhar, A., 1998. Chromosome numbers of *Tritrichomonas foetus* and *Tritrichomonas suis*. Vet. Parasitol. 78, 247–251.
Yao, C., 2013. Diagnosis of *Tritrichomonas foetus*-infected bulls, an ultimate approach to eradicate bovine trichomoniasis in US cattle?. J. Med. Microbiol. 62, 1–9.
Zubácová, Z., Cimbůrek, Z., Tachezy, J., 2008. Comparative analysis of trichomonad genome sizes and karyotypes. Mol. Biochem. Parasitol. 161, 49–54.