Parasitology Research (2019) 118:355–361
https://doi.org/10.1007/s00436-018-6161-7

PROTOZOOLOGY - SHORT COMMUNICATION

Clone-based haplotyping of *Giardia intestinalis* assemblage B human isolates

Lenka Lecová[1] · Pavla Tůmová[1] · Eva Nohýnková[1]

Received: 29 August 2018 / Accepted: 22 November 2018 / Published online: 28 November 2018

Abstract

The level of genetic variability of *Giardia intestinalis* clinical isolates is an intensively studied and discussed issue within the scientific community. Our collection of *G. intestinalis* human isolates includes six in vitro-cultured isolates from assemblage B, with extensive genetic variability. Such variability prevents the precise genotype characterisation by the multi-locus genotyping (MLG) method commonly used for assemblage A. It was speculated that the intra-assemblage variations represent a reciprocal genetic exchange or true mixed infection. Thus, we analysed gene sequences of the molecular clones of the assemblage B isolates, each representing a single DNA molecule (haplotype) to determine whether the polymorphisms are present within individual haplotypes. Our results, which are based on the analysis of three standard genetic markers (*bg*, *gdh*, *tpi*), point to haplotype diversity and show numerous single nucleotide polymorphisms (SNPs) mostly in codon wobble positions. We do not support the recombinatory origin of the detected haplotypes. The point mutations tolerated by mismatch repair are the possible cause for the detected sequence divergence. The precise sub-genotyping of assemblage B will require finding more conservative genes, as the existing ones are hypervariable in most isolates and prevent their molecular and epidemiological characterisation.

Keywords Assemblage B · *Giardia intestinalis* · Genetic variability · Haplotypes · Molecular cloning

Introduction

Giardia intestinalis (syn. *G. duodenalis*) is currently divided into eight genetically distinct groups called assemblages (designated A to H). The assemblages are morphologically indistinguishable and some of them host-specific. So-called zoonotic giardiasis is dominantly caused by two assemblages, A and B, with a relatively wide range of hosts, including humans. These two assemblages are further subdivided into sub-assemblages AI, AII, AIII, BIII, and BIV (Sprong et al. 2009), representing clusters of genetically close, but not identical, isolates within each assemblage (Ryan and Cacciò 2013). Determination of the *Giardia* genetic groups is based on sequence analysis of variable genetic loci. Currently for this purpose, three genetic markers (β-giardin (*bg*), glutamate dehydrogenase (*gdh*), and triose phosphate isomerase (*tpi*)) with a high, though variable, degree of genetic polymorphism are the most used. Nevertheless, genotyping analyses of particular assemblage B isolates are still limited due to the high genetic diversity encountered both between and within the isolates. In this assemblage, it is not possible to precisely identify a sub-assemblage and its subtypes (MLG), as it is routinely used for assemblage A. It is assumed that this variability is recorded due to the "real" mixed (sub-) assemblage infection, allelic sequence heterozygosity (ASH), or potentially a mixture of these two factors (Cacciò et al. 2008; Franzèn et al. 2009; Lebbad et al. 2011; Wielinga et al. 2011; Ankarklev et al. 2012; Faria et al. 2017). A combination of both of these situations was also confirmed at the single-cell level—in single trophozoites and in single cysts from clinical samples (Ankarklev et al. 2012). Moreover, these studies also assume genetic recombination both between and/or within assemblages as an alternative explanation for the abovementioned genetic diversity. In addition to biological cloning, which provides information from a single cell, polymorphisms in assemblage B human isolates are also investigated by molecular cloning. This approach (sequencing of multiple plasmids

Section Editor: Lihua Xiao

✉ Lenka Lecová
lenka.lecova@lf1.cuni.cz

[1] Institute of Immunology and Microbiology, First Faculty of Medicine, Charles University, Studničkova 7, 128 00 Prague, Czech Republic

containing cloned PCR products) provides information from a single DNA molecule. Some studies (Lasek-Nesselquist et al. 2009; Kosuwin et al. 2010; Siripattanapipong et al. 2011) identified a greater number of haplotypes in single isolates and point to putative genetic recombination. In our study, the same molecular approach was used to determine the extent and nature of genetic variability in our clinical isolates, which would allow for a better understanding of the genotyping failure in the assemblage B samples, a common problem in *Giardia* diagnostics and epidemiology.

Material and methods

Five human assemblage B clinical isolates (VZ-15, DM-16, AP-15, 49, and HH) described in our previous study (Lecová et al. 2018) were used.

All primary sequences were obtained from direct sequencing of PCR products amplified from genomic DNA that was extracted from in vitro cultures of the isolates. The PCR protocol described in detail in Lecová et al. (2018) was used with two modifications as follows: Q5® High-Fidelity DNA Polymerase was used for DNA amplification instead of *Taq* DNA Polymerase due to its ultra-low error rates (100× lower than *Taq* DNA Polymerase) and inosine base in both primers for the *tpi* gene was replaced by cytosine (I > C) due to its incompatibility with high-fidelity polymerases. All PCRs included an initial denaturation step at 98 °C for 30 s, followed by 30 cycles, each including denaturation at 98 °C for 10 s, annealing for 20 s (*bg* gene at 66 °C, *gdh* gene at 55 °C, and *tpi* gene at 60 °C), extension at 72 °C for 20 s, and final extension at 72 °C for 2 min. All PCRs were prepared in a total volume of 25 μL containing 0.25 units of Q5® High-Fidelity DNA Polymerase, 1× Q5 Reaction Buffer, 1× Q5 High GC Enhancer (New England Biolabs, Ipswich, MA, USA), 200 μM of 2′-deoxynucleoside 5′-triphosphates (dNTP Mix, PCR-Grade) (Qiagen, Hilden, Germany), 0.5 μM of each primer, nuclease-free water, and 1 μL of genomic DNA. All PCR products were purified using the Gel/PCR DNA Fragments Extraction Kit (Geneaid, New Taipei City, Taiwan) and subsequently sent to the DNA Sequencing Laboratory in the core facilities of the Faculty of Science, Charles University in Prague with the respective forward and reverse primers. Sequences were edited and aligned by using the sequence analysis software Geneious (Version 10.2.3, Biomatters Ltd., Auckland, New Zealand) and all chromatograms were inspected visually. If any polymorphism (compared to the reference gene sequence) or a double signal in the chromatograms was detected, cloning of this PCR product was performed to confirm the existence of multiple templates.

For cloning, the purified PCR products were ligated into pJET 1.2/blunt cloning vectors (CloneJET PCR Cloning Kit, Thermo Fisher Scientific, Waltham, MA, USA) and subsequently introduced into *Escherichia coli* TOP10 competent cells by heat-shock transformation. The recombinant plasmids were purified from 20 positive clones of each sample and gene by the High-Speed Plasmid Mini Kit (Geneaid) and sequenced using pJET1.2 forward and reverse sequencing primers. All obtained haplotype sequences were aligned and compared to sequences of the original PCR products.

All detected nonsynonymous substitutions were analysed using PROVEAN (Protein Variation Effect Analyser) v1.1 (Choi et al. 2012) to predict whether protein sequence variations affect protein function. A default cut-off value of − 2.5 (where a mutation is considered deleterious at − 2.5 and a higher negative value predicts a higher deleterious effect) was used. PROVEAN introduces a delta alignment score based on the reference and variant versions of a protein query sequence with respect to sequence homologues collected from the NCBI NR protein database through BLAST.

Results and discussion

Our previous study (Lecová et al. 2018) confirmed assemblage B in six isolates, but almost all sequences could not be unequivocally assigned to sub-assemblages or subtypes. A similar unclear subgrouping within assemblage B has already been reported (e.g., Robertson et al. 2006; Feng and Xiao 2011). Due to the variable levels of intra-isolate sequence heterozygosity and/or the numerous SNPs at the sub-assemblage-specific sites, we decided to analyse these samples in detail by haplotyping based on molecular cloning and sequencing. For this purpose, we selected different combinations of five isolates and three genes with SNPs and/or heterozygous positions compared to the reference genes (see, Tables 1, 2, and 3). Samples without polymorphic gene sequences, i.e., samples identical to the reference genes, were not cloned (VZ-15 *gdh*, VZ-15 *tpi*, and 49 *tpi*). Interestingly, in some samples, using the Q5® High-Fidelity DNA Polymerase caused a loss of all heterozygous positions (AP-15 *bg*, *tpi*, and HH *tpi*) compared to our previous study where *Taq* polymerase was used. Nevertheless, we cloned these samples, and two variants at the heterozygous sites were found among the haplotypes.

All cloning and sequencing results are summarised in Tables 1, 2, 3, and 4. (i) Within the *bg* gene, one to three different haplotype sequences per sample were found (Tables 1 and 4). The examined fragment of the *bg* gene (approx. 450 bp) has seven positions specific for the differentiation of sub-assemblages BIII and BIV. Table 4 also shows the ratios of these specific positions (BIII:BIV) found in individual haplotypes. This positions-ratio in the *bg* gene for samples VZ-15 and 49 corresponds to their classification to sub-assemblage BIV and the ratio in other samples (DM-16, AP-15, and HH) classified as sub-assemblage BIII. Since no

Table 1 SNPs and heterozygous positions in the *bg* gene among *Giardia intestinalis* assemblage B human isolates and their haplotypes

Sample ID (GenBank acc. no.)	Nucleotide position, *bg*																
	204	*210*	216	222	228	*273*	301	*327*	*354*	*357*	369	378	*438*	516	564	609	*648*
BIII (AY072726)	g	c[1)]	a	c	a	g[1)]	c	c[1)]	c[1)]	c[1)]	c	c	c[1)]	t	t	c	g[1)]
BIV (AY072725)	.	T[2)]	.	.	.	A[2)]	.	T[2)]	T[2)]	T[2)]	.	.	T[2)]	.	.	.	A[2)]
VZ-15 (MG558336)	A	T[2)]	.	.	.	A[2)]	.	.[1)]	T[2)]	T[2)]	T	.	T[2)]	C	.	.	.[1)]
Haplotype 1–20	A	T[2)]	.	.	.	A[2)]	.	.[1)]	T[2)]	T[2)]	T	.	T[2)]	C	.	.	.[1)]
DM-16 (MG558337)	.	.[1)]	R	.	G	A[2)]	.	.[1)]	T[2)]	T[2)]	.	.	.[1)]	.	C	T	.[1)]
Haplotype 1–11	.	.[1)]	.	.	G	A[2)]	.	.[1)]	T[2)]	T[2)]	.	.	.[1)]	.	C	T	.[1)]
Haplotype 12–20	.	.[1)]	G	.	G	A[2)]	.	.[1)]	T[2)]	T[2)]	.	.	.[1)]	.	C	T	.[1)]
AP-15 (MG558338)	.	.[1)]	.	.	G	A[2)]	T	.[1)]	.[1)]	T[2)]	.	T	.[1)]	.	.	T	.[1)]
Haplotype 1–18	.	.[1)]	.	.	G	A[2)]	T	.[1)]	.[1)]	T[2)]	.	T	.[1)]	.	.	T	.[1)]
Haplotype 19	.	.[1)]	.	.	.	A[2)]	.	.[1)]	.[1)]	T[2)]	.	.	.[1)]	.	.	.	.[1)]
Haplotype 20	.	.[1)]	.	.	.	A[2)]	.	.[1)]	.[1)]	T[2)]	.	.	.[1)]	.	C	T	.[1)]
49 (MG558339)	.	T[2)]	.	.	.	A[2)]	.	.[1)]	T[2)]	T[2)]	.	.	T[2)]	.	.	.	.[1)]
Haplotype 1–20	.	T[2)]	.	.	.	A[2)]	.	.[1)]	T[2)]	T[2)]	.	.	T[2)]	.	.	.	.[1)]
HH (MG558340)	.	.[1)]	.	Y	.	A[2)]	.	.[1)]	Y	T[2)]	.	.	.[1)]	.	.	.	.[1)]
Haplotype 1–11	.	.[1)]	.	.	.	A[2)]	.	.[1)]	.[1)]	T[2)]	.	.	.[1)]	.	.	.	.[1)]
Haplotype 12–20	.	.[1)]	.	T	.	A[2)]	.	.[1)]	T[2)]	T[2)]	.	.	.[1)]	.	.	.	.[1)]
Amino acid	A	L	R	I	T	E	L	T	L	A	D	N	L	Y	A	A	E

Nucleotide substitutions (capital letters) are numbered from ATG codon of the *bg* gene, dots indicate identity to the BIII (GenBank Accession No. AY072726) reference sequence, and heterozygous positions are indicated by standard IUPAC codes. The sub-assemblage-specific positions are labelled in italics. The haplotypes correspond to the genotype of cloned samples and reveal the combination of both BIII[1)] and BIV[2)] sub-assemblages

haplotype in the *bg* gene was the same as the reference sub-assemblage sequence, our results indicate that these isolates underwent either genetic recombination between both sub-assemblages or sequence differentiation by point mutations at the sub-assemblage-specific sites. (ii) The *gdh* gene fragment (approx. 650 bp) was no longer unambiguous. In the *gdh* gene, ten sub-assemblages-specific positions and one to eight different haplotypes per sample were detected (Tables 2 and 4). Samples DM-16 and 49 have a positions-ratio that corresponds to their classification into sub-assemblage BIII and BIV, respectively. However, sample AP-15 appears to belong to sub-assemblage BIV, inconsistent with the classification in the other two genes. The most controversial is sample HH, which showed three different haplotypes, one of which corresponds most likely to sub-assemblage BIV and the other two to sub-assemblage BIII. A similar co-presence of individual BIII and BIV haplotypes within a single sample was observed in Tůmová et al. (2018). (iii) In the *tpi* gene fragment (approx. 450 bp), five sub-assemblage-specific positions and one to three different haplotypes per sample were detected. All three investigated samples (DM-16, AP-15, and HH) have a ratio in these positions favouring the classification to sub-assemblage BIII (Tables 3 and 4).

In total, the sequence alignment in the *bg* gene revealed nine different haplotypes containing 15 synonymous substitutions (Table 1) compared to the reference gene BIV (Acc. No. AY072725). The nucleotide substitutions among 13 haplotypes of the *gdh* gene led to 24 synonymous changes and only one nonsynonymous change (Table 2) compared to the reference gene BIV (Acc. No. AY178738). The analysis of amino acid (AA) changes in six identified haplotypes of the *tpi* gene shows eight synonymous and four nonsynonymous substitutions (Table 3) compared with the BIV reference sequence (Acc. No. AF069560). These nonsynonymous changes, however, should not affect protein characteristics, as charged AA are replaced with another charged AA (K221R and K61R), polar AA for another polar AA (Y31H and H91Y), and hydrophobic AA for another hydrophobic AA (I102V). Moreover, to verify function retention of the modified protein, all nonsynonymous substitutions were analysed by PROVEAN v1.1, and the results are displayed in Table 5. All five nonsynonymous AA substitutions were predicted to be neutral (non-deleterious).

Our results show that the intra-isolate clonal diversity—expressed in the number of identified haplotypes and SNPs in a haplotype per sample in assemblage B—is not as extensive as described in other studies (Lasek-Nesselquist et al. 2009; Kosuwin et al. 2010; Siripattanapipong et al. 2011). This result can be partially explained by the use of different DNA polymerases in the current study and other studies. High-fidelity amplification is essential for experiments whose outcome depends upon the correct DNA sequence. The

Table 2 SNPs and heterozygous positions in the *gdh* gene among *Giardia intestinalis* assemblage B human isolates and their haplotypes

Sample ID (GenBank acc. no.)	Nucleotide position, *gdh*																								
	219	237	*309*	357	360	*429*	*447*	486	516	519	*540*	546	*561*	570	597	606	*612*	662	690	*699*	705	*723*	756	*825*	834	867
BIII (AF069059)	t[1)]	c	c[1)]	t	g	t[1)]	t[1)]	t	g	c	c [1)]	c	c[1)]	c	c	c	g[1)]	a	g	t[1)]	t	c[1)]	t	a[1)]	c	c
BIV (AY178738)	C[2)]	.	T[2)]	.	.	C[2)]	C[2)]	.	.	.	T[2)]	.	T[2)]	.	.	.	A[2)]	.	.	C[2)]	.	T[2)]	.	G[2)]	.	.
DM-16 (MG558341)	.[1)]	Y	T[2)]	Y	R	.[1)]	.[1)]	Y	.	T	.[1)]	Y	.[1)]	.	.	.	.[1)]	.	.	C[2)]	.	.[1)]	.	.[1)]	T	Y
Haplotype 1–5	.[1)]	T	T[2)]	.	.	.[1)]	.[1)]	.	.	T	.[1)]	.	.[1)]	.	.	.	. [1)]	.	.	C[2)]	.	.[1)]	.	.[1)]	T	.
Haplotype 6–14	.[1)]	T	T[2)]	.	.	.[1)]	.[1)]	.	.	T	.[1)]	.	.[1)]	.	.	.	.[1)]	.	.	C[2)]	.	.[1)]	.	.[1)]	T	T
Haplotype 15	.[1)]	T	T[2)]	C	A	.[1)]	.[1)]	C	.	T	.[1)]	T	.[1)]	.	.	.	.[1)]	.	.	C[2)]	.	.[1)]	.	.[1)]	T	.
Haplotype 16	.[1)]	.	T[2)]	C	A	.[1)]	.[1)]	C	.	T	.[1)]	T	.[1)]	.	.	.	.[1)]	.	.	C[2)]	.	.[1)]	.	.[1)]	T	.
Haplotype 17	.[1)]	T	T[2)]	.	.	.[1)]	.[1)]	C	.	T	.[1)]	T	.[1)]	.	.	.	.[1)]	.	.	C[2)]	.	.[1)]	.	.[1)]	T	.
Haplotype 18	.[1)]	.	T[2)]	.	.	.[1)]	.[1)]	.	.	T	.[1)]	.	.[1)]	.	.	.	.[1)]	.	.	C[2)]	.	.[1)]	.	.[1)]	T	T
Haplotype 19	.[1)]	T	T[2)]	C	A	.[1)]	.[1)]	C	.	T	.[1)]	T	.[1)]	.	.	.	.[1)]	.	.	C[2)]	.	.[1)]	.	.[1)]	T	T
Haplotype 20	.[1)]	T	T[2)]	C	A	.[1)]	.[1)]	.	.	T	.[1)]	.	.[1)]	.	.	.	.[1)]	.	.	C[2)]	.	.[1)]	.	.[1)]	T	T
AP-15 (MG558342)	C[2)]	.	T[2)]	C	.	.[1)]	C[2)]	.	.	.	T[2)]	.	.[1)]	.	.	.	A[2)]	.	.	C[2)]	.	T[2)]	.	.[1)]	T	.
Haplotype 1–20	C[2)]	.	T[2)]	C	.	.[1)]	C[2)]	.	.	.	T[2)]	.	.[1)]	.	.	.	A[2)]	.	.	C[2)]	.	T[2)]	.	.[1)]	T	.
49 (MG558343)	C[2)]	.	T[2)]	C	.	C[2)]	C[2)]	.		.	T[2)]	.	.[1)]	T	T	.	A[2)]	.	A	C[2)]	C	T[2)]	C	G[2)]	.	.
Haplotype 1–20	C[2)]	.	T[2)]	C	.	C[2)]	C[2)]	.	.	.	T[2)]	.	.[1)]	T	T	.	A[2)]	.	A	C[2)]	C	T[2)]	C	G[2)]	.	.
HH (MG558344)	.[1)]	.	.[1)]	Y	R	Y	Y	Y	R	T	.[1)]	Y	.[1)]	.	.	Y	R	R	.	C[2)]	.	Y	.	R	Y	Y
Haplotype 1–8	.[1)]	.	.[1)]	C	A	C[2)]	C[2)]	C	.	T	.[1)]	.	.[1)]	.	.	T	A[2)]	G	.	C[2)]	.	T[2)]	.	G[2)]	.	.
Haplotype 9–19	.[1)]	.	.[1)]	.	.	.[1)]	.[1)]	.	A	T	.[1)]	T	.[1)]	.	.	.	.[1)]	.	.	C[2)]	.	.[1)]	.	.[1)]	T	T
Haplotype 20	.[1)]	.	.[1)]	.	.	.[1)]	.[1)]	.	A	T	.[1)]	T	.[1)]	.	.	T	A[2)]	G	.	C[2)]	.	T[2)]	.	.[1)]	T	T
Amino acid	V	I	L	L	P	F	T	P	E	I	Y	R	F	V	G	L	R	K221R	K	L	S	A	G	K	H	R

Nucleotide substitutions (capital letters) are numbered from ATG codon of the *gdh* gene, dots indicate identity to the BIII (GenBank Accession No. AF069059) reference sequence, and heterozygous positions are indicated by standard IUPAC codes. The sub-assemblage-specific positions are labelled in italics. The haplotypes correspond to the genotype of cloned samples and reveal the combination of both BIII[1)] and BIV[2)] sub-assemblages. Amino acid substitutions are numbered from the translation initiator methionine (numbered as + 1)

Table 3 SNPs and heterozygous positions in the *tpi* gene among *Giardia intestinalis* assemblage B human isolates and their haplotypes

Sample ID (GenBank acc. no.)	Nucleotide position, *tpi*											
	91	162	*165*	*168*	182	*210*	271	304	342	393	*429*	537
BIII (AF069561)	c [1)]	g	c[1)]	c[1)]	a	g[1)]	c	a	g	c	g[1)]	g
BIV (AF069560)	T[2)]	.	T[2)]	T[2)]	.	A[2)]	.	.	.	.	A[2)]	.
DM-16 (MG558345)	.[1)]	.	.[1)]	.[1)]	.	.[1)]	.	R	.	Y	.[1)]	A
Haplotypes 1–19	.[1)]	.	.[1)]	.[1)]	.	.[1)]	.	.	.	.	.[1)]	A
Haplotype 20	.[1)]	.	.[1)]	.[1)]	.	.[1)]	.	G	.	T	.[1)]	A
AP-15 (MG558346)	.[1)]	A	.[1)]	.[1)]	.	.[1)]	.	.	A	.	.[1)]	.
Haplotypes 1–18	.[1)]	A	.[1)]	.[1)]	.	.[1)]	.	.	A	.	.[1)]	.
Haplotype 19	.[1)]	.	T[2)]	.[1)]	.	.[1)]	.	.	A	.	.[1)]	.
Haplotype 20	.[1)]	.	T[2)]	.[1)]	.	A[2)]	.	.	.	.	.[1)]	A
HH (MG558347)	.[1)]	.	T[2)]	T[2)]	G	.[1)]	T	.	.	.	.[1)]	.
Haplotypes 1–20	.[1)]	.	T[2)]	T[2)]	G	.[1)]	T	.	.	.	.[1)]	.
Amino acid	Y31H	A	N	T	K61R	E	H91Y	I102V	A	T	V	V

Nucleotide substitutions (capital letters) are numbered from ATG codon of the *tpi* gene, dots indicate identity to the BIII (GenBank Accession No. AF069561) reference sequence, and heterozygous positions are indicated by standard IUPAC codes. The sub-assemblage-specific positions are labelled in italics. The haplotypes correspond to the genotype of cloned samples and reveal the combination of both BIII[1)] and BIV[2)] sub-assemblages, except the sample DM-16 which corresponding to BIII sub-assemblage only. Amino acid substitutions are numbered from the translation initiator methionine (numbered as + 1)

abovementioned studies used a low-fidelity polymerase such as *Taq* DNA Polymerase (Lasek-Nesselquist et al. 2009) or its derivatives with a slightly higher fidelity Ex*Taq*® DNA Polymerase (Kosuwin et al. 2010) and Go*Taq*® DNA Polymerase (Siripattanapipong et al. 2011). These enzymes are characterised by a high error rate, and thus are often not explicitly recommended for cloning/subcloning from in vitro-amplified material (PCR, WGA, etc.). In our study, the highest fidelity DNA polymerase available was used (Q5® High-Fidelity DNA Polymerase). Its fidelity is 100× higher than *Taq*, while Ex*Taq*® fidelity is better than *Taq* by 4.5× (as reported by the suppliers). On the other hand, cloning and sequencing was also tried in our PCR products (*gdh* gene) amplified with the conventional *Taq* polymerase (Thermo Fisher Scientific). In this case, six to 23 additional SNPs (per ten clones of one sample) appeared compared to Q5® polymerase (data not shown).

We observed heterozygosity in 50% of our samples. The presence of double peaks from DNA sequencing within the assemblage B isolates is usually interpreted as genetic recombination or mixed infections. The combinations of both sub-assemblage BIII and BIV positions were recorded in the majority (84%) of haplotypes obtained in our study. This excludes the interpretation as a solely true mixed infection by BIII and BIV cells and points rather to recombination. The evidence of intra-assemblage B recombination was previously indirectly found and described based on statistical evaluation using recombination detection programmes (Kosuwin et al. 2010; Siripattanapipong et al. 2011). We did not confirm such an extent of heterogeneity as obtained by previous studies that used error-prone DNA polymerases. Moreover, we believe that the observed SNPs at BIII and BIV sub-assemblage-specific and other positions did not result from recombination but represent endogenous sequence variability. We base our

Table 4 Numerical overview of haplotype variants per sample and ratios of sub-assemblage-specific positions (BIII:BIV) in individual haplotypes

Sample ID	*bg* gene		*gdh* gene		*tpi* gene	
	No. of haplotype variants	Ratio of BIII:BIV positions	No. of haplotype variants	Ratio of BIII:BIV positions	No. of haplotype variants	Ratio of BIII:BIV positions
VZ-15	1	2:5	–	–	–	–
DM-16	2	4:3	8	8:2	2	5:0
AP-15	3	5:2	1	3:7	3*	5:0/4:1/3:2
49	1	2:5	1	1:9	–	–
HH	2*	5:2/4:3	3*	6:4/9:1/3:7	1	3:2

*Haplotype variants differing also in sub-assemblage-specific positions (more ratios detected)

Table 5 PROVEAN analysis of nonsynonymous substitutions

Gene	Variant	No. of sequences	No. of clusters	PROVEAN score	Prediction (cut-off = − 2.5)
gdh	K221R	517	30	− 1.079	Neutral
tpi	Y31H	776	30	0.976	Neutral
tpi	K61R	776	30	− 0.354	Neutral
tpi	H91Y	776	30	3.123	Neutral
tpi	I102V	776	30	− 0.487	Neutral

Amino acid substitutions are numbered from the translation initiator methionine (numbered as + 1)

assumption on the positions-ratio between BIII and BIV, which is indicative of a surprisingly high amount of BIII/BIV combinations on a single DNA molecule in a very short gene portion (e.g., five combinations within 450 bp in sample DM-16). This is probably not achieved by multiple bilateral recombination events, as the chiasma interference limits the crossover number and ensures that multiple crossovers along the chromosome are widely spaced (reviewed in Chakraborty et al. 2017). Most of the detected positions with SNPs including the sub-assemblage-specific positions are located in codon wobble positions (91%), which are fault-tolerant and prone to point mutations but do not cause AA substitutions. We thus believe that the observed sequence variability is not the product of homologous recombination. It rather refers to spontaneous silent mutations that are tolerated and fixed, although individual events of the polymorphous recombination process cannot be excluded. The mismatch repair (MMR) genes *MutS* and *MutL* in *Giardia* appear to be derived from their eukaryotic homologues (Lin et al. 2007). Moreover, some differences in MMR may be particularly present in assemblages B in *Giardia*, which have tenfold higher levels of allelic sequence heterozygosity (ASH) compared to assemblage A (Franzèn et al. 2009). The elevated ASH in assemblage B is attributed to the absence of homologous recombination (Morrison and Svärd 2011). Next-generation sequencing (NGS), due to the amount of the gained data, could also contribute to the validation of our hypothesis that the polymorphisms are caused by point mutations preferably in codon wobble positions.

Our study also indicates that intra-assemblage recombination/sequence divergence is likely to occur even when the primary gene sequences do not show heterozygous positions (double peaks). The hidden variability is not surprising in Sanger sequencing and must be taken into account in *Giardia* cultures. The clonal-based proliferation may favour some variants to overgrow the others under conditions of stress (Tibayrenc and Ayala 2014). However, the observed conservation in protein structures and function of the three investigated genes and their variants (haplotypes) in our study refers rather to the presence of *Giardia* endogenous mechanisms, which sort-out the deleterious mutations and maintain silent nucleotide variability.

Conclusion

Cloning and sequencing approaches were used to elucidate the heterologous alleles present within samples of *Giardia intestinalis* assemblage B clinical isolates. Our results, which are based on the analysis of three standard genetic markers (*bg*, *gdh*, *tpi*), point to intra-isolate clonal diversity and show numerous SNPs in codon wobble positions; which are silent nucleotide substitutions that do not lead to an amino acid change. In a few positions where the amino acid was substituted, such change did not affect the protein structure and function. We do not support the recombinatory origin of the detected haplotypes. All previously described results were indirectly assumed from mathematical and statistical models, although they contradict topological set-ups of recombination events. We hypothesise that the point mutations tolerated by MMR are one possible cause for the detected sequence divergence, as most SNPs are located in codon wobble positions. Thus, the following question remains: How appropriate is the use of these three particular genes for assembling the assemblage B clinical isolates due to the hypervariability in codon wobble positions? More conservative markers for assemblage B may be needed.

Financial support This study was supported by the Czech Health Research Council (Grant No. 15-33369A).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval For this type of study, formal consent is not required.

References

Ankarklev J, Svärd SG, Lebbad M (2012) Allelic sequence heterozygosity in single *Giardia* parasites. BMC Microbiol 12(65):1–10. https://doi.org/10.1186/1471-2180-12-65

Cacciò SM, Beck R, Lalle M, Marinculic A, Pozio E (2008) Multilocus genotyping of *Giardia duodenalis* reveals striking differences between assemblages A and B. Int J Parasitol 38(13):1523–1531. https://doi.org/10.1016/j.ijpara.2008.04.008

Chakraborty P, Pankajam AV, Lin G, Dutta A, Nandanan KG, Tekkedil MM, Shinohara A, Steinmetz LM, Thazath NK (2017) Modulating crossover frequency and interference for obligate crossovers in *Saccharomyces cerevisiae* meiosis. Genes Genomes Genet 7(5): 1511–1524. https://doi.org/10.1534/g3.117.040071

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and indels. PLoS One 7(10):e46688. https://doi.org/10.1371/journal.pone.0046688

Faria CP, Zanini GM, Dias GS, da Silva S, do Céu Sousa M (2017) New multilocus genotypes of *Giardia lamblia* human isolates. Infect Genet Evol 54:128–137. https://doi.org/10.1016/j.meegid.2017.06.028

Feng Y, Xiao L (2011) Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis. Clin Microbiol Rev 24(1):110–140. https://doi.org/10.1128/CMR.00033-10

Franzèn O, Jerlström-Hultqvist J, Castro E, Sherwood E, Ankarklev J, Reiner DS, Palm D, Andersson JO, Andersson B, Svärd SG (2009) Draft genome sequencing of *Giardia intestinalis* assemblage B isolate GS: is human giardiasis caused by two different species? PLoS Pathog 5(8):e1000560. https://doi.org/10.1371/journal.ppat.1000560

Kosuwin R, Putaporntip C, Pattanawong U, Jongwutiwes S (2010) Clonal diversity in *Giardia intestinalis* isolates from Thailand: evidences of intragenic recombination and purifying selection at the beta giardin locus. Gene 449:1–9. https://doi.org/10.1016/j.gene.2009.09.010

Lasek-Nesselquist E, Welch DM, Thompson RC, Steuart RF, Sogin ML (2009) Genetic exchange within and between assemblages of *Giardia duodenalis*. J Eukaryot Microbiol 56:504–518. https://doi.org/10.1111/j.1550-7408.2009.00443.x

Lebbad M, Petersson I, Karlsson L, Botero-Kleiven S, Andersson JO, Svenungsson B, Svärd SG (2011) Multilocus genotyping of human *Giardia* isolates suggests limited zoonotic transmission and association between assemblage B and flatulence in children. PLoS Negl Trop Dis 5(8):e1262. https://doi.org/10.1371/journal.pntd.0001262

Lecová L, Weisz F, Tůmová P, Tolarová V, Nohýnková E (2018) The first multilocus genotype analysis of *Giardia intestinalis* in humans in the Czech Republic. Parasitology 145(12):1577–1587. https://doi.org/10.1017/S0031182018000409

Lin Z, Nei M, Ma H (2007) The origins and early evolution of DNA mismatch repair genes - multiple horizontal gene transfers and co-evolution. Nucleic Acids Res 35(22):7591–7603. https://doi.org/10.1093/nar/gkm921

Morrison HG, Svärd S (2011) Genomics of *Giardia*. In: Luján HD, Svärd S (eds) Giardia a model organism. Springer, Wien, pp 95–101

Robertson LJ, Hermansen L, Gjerde BK, Strand E, Alvsvåg JO, Langeland N (2006) Application of genotyping during an extensive outbreak of waterborne giardiasis in Bergen, Norway, during autumn and winter 2004. Appl Environ Microbiol 72(3):2212–2217. https://doi.org/10.1128/AEM.72.3.2212-2217.2006

Ryan U, Cacciò SM (2013) Zoonotic potential of *Giardia*. Int J Parasitol 43(12–13):943–956. https://doi.org/10.1016/j.ijpara.2013.06.001

Siripattanapipong S, Leelayoova S, Mungthin M, Thompson RC, Boontanom P, Saksirisamphant W, Tan-Ariya P (2011) Clonal diversity of the glutamate dehydrogenase gene in *Giardia duodenalis* from Thai isolates: evidence of genetic exchange or Mixed Infections? BMC Microbiol 11:206. https://doi.org/10.1186/1471-2180-11-206

Sprong H, Cacciò SM, der Giessen JWB V, on behalf of the ZOOPNET network and partners (2009) Identification of zoonotic genotypes of *Giardia duodenalis*. PLoS Negl Trop Dis 3(12):e588. https://doi.org/10.1371/journal.pntd.0000558

Tibayrenc M, Ayala FJ (2014) *Cryptosporidium, Giardia, Cryptococcus, Pneumocystis* genetic variability: cryptic biological species or clonal near-clades? PLoS Pathog 10(4):e1003908. https://doi.org/10.1371/journal.ppat.1003908

Tůmová P, Mazánek L, Lecová L, Dluhošová J, Typovská H, Kotrašová V, Ticháčková V, Nohýnková E (2018) A natural zoonotic giardiasis: infection of a child via *Giardia* cysts in pet chinchilla droppings. Parasitol Int 67(6):759–762. https://doi.org/10.1016/j.parint.2018.07.010

Wielinga C, Ryan U, Andrew Thompson RC, Monis P (2011) Multilocus analysis of *Giardia duodenalis* intra-Assemblage B substitution patterns in cloned culture isolates suggests sub-Assemblage B analyses will require multi-locus genotyping with conserved and variable genes. Int J Parasitol 41(5):495–503. https://doi.org/10.1016/j.ijpara.2010.11.007