G-quadruplexes in helminth parasites

Alessio Cantara[1,2,†], Yu Luo[3,4,†], Michaela Dobrovolná[5,†], Natalia Bohalova[1,2], Miroslav Fojta[1], Daniela Verga [3,6], Lionel Guittat[4,7], Anne Cucchiarini[4], Solène Savrimoutou[8], Cécile Häberli[9,10], Jean Guillon[8], Jennifer Keiser[9,10], Václav Brázda[1,5,*] and Jean Louis Mergny [1,4,*]

[1]Institute of Biophysics, Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic, [2]Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5, 62500 Brno, Czech Republic, [3]CNRS UMR9187, INSERM U1196, Université Paris-Saclay, F-91405 Orsay, France, [4]Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, INSERM, Institut Polytechnique de Paris, 91128 Palaiseau, France, [5]Faculty of Chemistry, Brno University of Technology, Purkyňova 118, 612 00 Brno, Czech Republic, [6]CNRS UMR9187, INSERM U1196, Institut Curie, PSL Research University, F-91405 Orsay, France, [7]Université Sorbonne Paris Nord, UFR SMBH, Bobigny, France, [8]ARNA Laboratory, Université de Bordeaux, INSERM U1212, CNRS UMR 5320, UFR des Sciences Pharmaceutiques, Bordeaux, France, [9]Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland and [10]University of Basel, Basel, Switzerland

ABSTRACT

Parasitic helminths infecting humans are highly prevalent infecting ∼2 billion people worldwide, causing inflammatory responses, malnutrition and anemia that are the primary cause of morbidity. In addition, helminth infections of cattle have a significant economic impact on livestock production, milk yield and fertility. The etiological agents of helminth infections are mainly Nematodes (roundworms) and Platyhelminths (flatworms). G-quadruplexes (G4) are unusual nucleic acid structures formed by G-rich sequences that can be recognized by specific G4 ligands. Here we used the G4Hunter Web Tool to identify and compare potential G4 sequences (PQS) in the nuclear and mitochondrial genomes of various helminths to identify G4 ligand targets. PQS are nonrandomly distributed in these genomes and often located in the proximity of genes. Unexpectedly, a Nematode, *Ascaris lumbricoides,* was found to be highly enriched in stable PQS. This species can tolerate high-stability G4 structures, which are not counter selected at all, in stark contrast to most other species. We experimentally confirmed G4 formation for sequences found in four different parasitic helminths. Small molecules able to selectively recognize G4 were found to bind to *Schistosoma mansoni* G4 motifs. Two of these ligands demonstrated potent activity both against larval and adult stages of this parasite.

INTRODUCTION

Helminth infections caused by parasitic Nematodes (roundworms) and Platyhelminths (flatworms) are among the most prevalent afflictions for people living in poor areas of the world with over a quarter of the total human population affected worldwide (1,2). Parasitic worm infections cause anemia, malnutrition, allergies, bloody diarrhea, bowel cramps and inflammation associated with colonic polyposis (3) and increased susceptibility to HIV and progression to AIDS, resulting in many obstructive pathologies (3,4). In addition, severe anemia in pregnancy is associated with neonatal prematurity and reduced birthweight (3,5). Cattle infections have a significant economic impact on livestock production due to a reduction in growth, milk yield and fertility (6). Recent estimates suggest that *Ascaris lumbricoides* infects over a billion, *Trichuris trichiura* 795 million (7), and *Strongyloides stercoralis* 30–100 million people (8). Hookworms such as *Necator americanus* and *Ancylostoma duodenale* cause hookworm diseases, which are associated with blood loss and anemia. Schistosomiasis is the third most reported global tropical disease caused by trematode flukes of the genus Schistosoma (9). The three most common species of parasitic trematodes of the family Schistosomatidae are *Schistosoma mansoni, Schistosoma japonicum* and *Schistosoma haematobium* (1,9). The greatest numbers of *S. mansoni* infections occur in Sub-Saharan

Africa, the Middle East, the Caribbean, Brazil, Venezuela and Suriname. *S. japonicum* is localized to Asia, primarily the Philippines, Indonesia, and China. *S. haematobium* is highly prevalent in Sub-Saharan Africa, the Middle East, and was recently reported in Corsica (France) (2). Schistosomiasis is estimated to affect ~200 million people worldwide (10). Few treatments are available to fight worm infections. In addition, widespread use of drugs such as benzimidazoles selects for drug-resistant parasite strains (11), and reduced praziquantel efficacy leads to low egg reduction rates (12). The recent sequencing of the genome of helminths (*e.g. S. mansoni*, *S. japonicum* and *S. haematobium* (13–15) offers new opportunities to identify novel targets in key genes of the parasites.

Identifying potential G4-forming sequences (PQS) in key genes associated with the infectivity of parasitic worms should allow the prediction of new druggable targets for G4 ligands. G4 are noncanonical nucleic acid secondary structures (16) formed by G-rich sequences that are built of stacked tetrads (also called 'G-quartets') constituted of four Hoogsteen hydrogen-bonded guanine bases. They can display a wide variety of topologies, resulting from various possible combinations of DNA/RNA strand directions, as well as variations in loop size and molecularity (17). G4 are further stabilized by the presence of monovalent cations such as potassium or sodium, which are abundant in cells (18). G4 can be found throughout a variety of genomes (19), and are overrepresented in the promoter regions of genes, especially regulatory genes involved in cell proliferation or survival (20,21) as well as in regions which regulate important biological processes (22) including immune response, transcriptional activation, DNA damage repair (23) and telomere maintenance (24), as shown mostly in mammalian cells. The interest in the therapeutic potential of gene promoters containing G4 has resulted in a rapidly increasing number of studies during the past decade in which small molecules have been used to act as G4 stabilizers, with reports of transcription inhibition in cell-based assays.

G4 structures have only been investigated in a few Platyhelminths and Nematode species. For instance, 1,500 PQS were found in *Caenorhabditis elegans* using the *Quadparser* algorithm, of which ~500 are associated with transcription start site regions. The peak in PQS density coincides almost exactly with the nucleosome depleted region, which is consistent with the hypothesis that functional PQS may be located outside nucleosome bound regions (25). G4 stable under physiological conditions (K$^+$-rich buffer) and stabilized by Pyridostatin (PDS, a specific G4-targeting small molecule) were identified in *C. elegans* (26). The presence of G4 structures in heterochromatin and the difference in G4 staining between somatic and stem cells in germline DNA of the flatworm *Macrostomum lignano* pointed out the possibility that the resolution or suppression of G4 structures is important for stem cells with regenerative potential (27). A study conducted on *A. lumbricoides* shows that the fluorescent G4 ligand **Q1** binds selectively to the antiparallel telomeric G4 (28). Very recently, G4 motifs were identified in the genome of *S. mansoni* and the authors confirmed the presence of G4 in adult worms, by means of the BG4 G-quadruplex-specific antibody (29).

Given the lack of information available in regard to the nature of G4s formed in infective species of Platyhelminths and Nematodes, we analyzed differences in potential G4-forming sequences (PQS) presence, frequency and localization in the genomes of six Platyhelminths and four Nematodes using the G4Hunter tool. Comparing the genomes of parasites from these two phyla with different infectivity (parasitic and non-parasitic worms) may reveal details of processes driving pathogenicity and new possible drug targets. Finally, we showed that the recently-identified G4 motifs found in *S. mansoni* (29) may be targeted by G4 ligands, leading to anti-parasitic effects.

MATERIALS AND METHODS

Source of DNA sequences

The set of selected genomes (both mitochondrial and nuclear DNAs) of two Clades (Platyhelminths and Nematodes) belonging to the Clade Nephozoa (subgroup Protostomia) was downloaded from the Genome database of the NCBI. Six Platyhelminths (*Schistosoma haematobium*, *Schistosoma japonicum*, *Schistosoma mansoni*, *Trichobilharzia regenti*, *Dibothriocephalus latus* and *Taenia asiatica*), and four Nematoda (*Strongyloides stercoralis*, *Trichuris trichiura*, *Ascaris lumbricoides* and *Caenorhabditis elegans*) were chosen for an initial analysis. Four additional nematode species (*Anisakis simplex, Ascaris suum, Parascaris equorum* and *Toxocara canis)* were added for comparison with *Ascaris lumbricoides*. All accession numbers are provided in Supplementary Material 01.

Analytical process

All sequences were analyzed using the G4Hunter Web tool (30) (http://bioinformatics.ibp.cz/#/analyse/quadruplex) which can read National Center for Biotechnology Information (NCBI) IDs. Unless specified otherwise, parameters for G4Hunter were set to 25 nucleotides for window size and ≥1.2 for threshold scores. This threshold appears as a reasonable compromise, giving few false positives (sequences not forming a G4 despite a G4Hunter score above threshold) and false negatives (sequences able to form a stable G4 despite having a G4Hunter score below threshold). Scores above 1.2 correspond to sequences having a higher guanine content and likely to form stable G4s. To rank sequences based on score, motifs were binned in five intervals covering the G4Hunter scores 1.2–1.4, 1.4–1.6, 1.6–1.8, 1.8–2.0 and >2.0. Data were merged in a single Excel file where was made the statistical evaluation and are accessible in Supplementary Material 02.

To test whether the PQS occurrence in chromosomal breakpoint regions is significantly different than in a randomly shuffled sequence, we generated 40 random sequences with length and nucleotide content the same as in original *A. suum* breakpoints. All sequences were randomly shuffled using the Sequence Manipulation Suite (https://www.bioinformatics.org/sms2/shuffle_dna.html) and were manually analyzed using the G4Hunter Web tool (30). The parameters for G4Hunter were set to 25 nucleotides for window size and thresholds of 1.2 or more (several thresholds

were considered). Sequences were merged in a single Excel file where was made the statistical evaluation using *t*-test.

LOGO sequences

NCBI sequences in FASTA format were downloaded and the dataset was uploaded to SnapGene program. For every PQS we used the corresponding sequences from all analyzed genomes and alignments were generated using the Clustal Omega tool. All PQS found were searched in aligned sequences and WebLogo 3 was used for generating LOGO sequences Supplementary Material 04.

Statistical evaluation

Raw data were converted in .xlsx file format and analyzed through Microsoft Excel. All data files are available in Supplementary Materials. Correlation was evaluated by the Spearman's rank correlation coefficient (r_s) and are presented in Supplementary Material 05.

Samples

Oligonucleotides were purchased from Eurogentec (Belgium) and used without further purification. Stock solutions were prepared at 100 μM strand concentration for the unlabeled oligonucleotides and at 200 μM strand concentration for double-labeled oligonucleotides in ddH$_2$O. Sequences of tested G4 motifs and control G4s, single-strands, and duplexes are shown in Supplementary Table S5. All oligonucleotides were annealed (95°C for 5 min and slowly cooled to room temperature) in the corresponding buffer before measurements.

UV-melting assay

3 μM oligonucleotide solutions were annealed in K100 buffer (100 mM KCl, 10 mM lithium cacodylate, 90 mM LiCl, pH 7.2). UV-melting profiles were recorded with a Cary 300 spectrophotometer (Agilent Technologies, France). Heating runs were performed between 10°C and 95°C, the temperature was increased by 0.2°C/min, and absorbance was recorded at 260 and 295 nm (31).

Circular dichroism

3 μM oligonucleotide solutions were annealed in K100 buffer. CD spectra were recorded on a J-1500 spectropolarimeter (JASCO, France) at room temperature (25°C), using a scan range of 300–220 nm, a scan rate of 100 nm/min and averaging four accumulations.

FRET-melting assay

FRET melting assay was performed in 96-well plates and the fluorescence of dual-labeled G4-forming oligonucleotides (including F21T; sequences are shown in Supplementary Table S7) was recorded using a CFX96 qPCR instrument (Biorad). F21T sequence was annealed at 0.23 μM in K10 buffer (10 mM KCl, 10 mM lithium cacodylate, 90 mM LiCl, pH 7.2), then the oligonucleotide was added to each well (final strand concentration of 0.2 μM) which was incubated with or without the tested ligands at 2 and 5 μM final concentration, to a final volume of 25 μl. Competition experiments were performed in the presence of non-labeled sequences, including one auto complementary duplex, ds26, one parallel G4, c-myc and G4s from *S. Mansoni*. The microplate was incubated at 25°C for 5 min, after which the temperature was increased by increments of 0.5°C/min to reach 95°C. The collected signal was normalized to 1 and the melting temperature (T_m) was defined when the normalized signal was 0.5. ΔT_m corresponds to the difference of T_m between the oligonucleotides with and without the ligands. This FRET-melting assay was done in duplicates.

FRET-melting competition assay

FRET melting competition (FRET-MC) experiments were performed in 96-well plates using a HT7900 RT-PCR instrument (Applied BioSystem), as previously described (32). 50 μM oligonucleotide solutions were annealed in K10 buffer. Each well contained 3 μM competitors, 0.2 μM fluorescent oligonucleotide F21T in the presence or absence of 0.4 μM G4 ligand (PhenDC3) in K10 buffer, for a final volume of 25 μl. Samples were kept at 25°C for 5 min, then the temperature was increased by 0.5°C per minute until 95°C, and the FAM channel was used to collect the fluorescence signal. The T_m of an oligonucleotide is defined as the temperature at which 50% of the oligonucleotide is unfolded. ΔT_m is determined as the difference in T_m with the sample containing F21T in the absence of PhenDC3. Each experimental condition was tested in duplicate on two separate plates.

Thermal difference spectra (TDS) and Isothermal difference spectra (IDS)

Absorbance spectra were recorded on a Cary 300 spectrophotometer (Agilent Technologies, France) (scan range: 500–200 nm; scan rate: 600 nm/min; automatic baseline correction).

- **TDS**: 3 μM oligonucleotide solutions were annealed in K100 buffer. After recording the first spectra (folded) at 25°C, temperature was increased to 95°C, and the second UV-absorbance spectra was recorded after 15 min of equilibration at high temperature. TDS corresponds to the arithmetic difference between the initial (folded; 25°C) and second (unfolded; 95°C) spectra (33).
- **IDS**: 3 μM oligonucleotide solutions were annealed in Li-Caco10 buffer (10 mM lithium cacodylate, pH 7.2). Absorbance spectra were first recorded at 25°C in the absence of any stabilizing cation. 1 M KCl was added after recording the first spectrum, to a final potassium concentration of 100 mM KCl. The second UV-absorbance spectrum was recorded after 15 min of equilibration. IDS correspond to the arithmetic difference between the initial (unfolded) and final (folded, thanks to the addition of K$^+$) spectra, after correction for dilution.

G-Quadruplex fluorescent light-up probes

- **ThT** (Thioflavin T) was used as previously described (34). 7.5 μM oligonucleotide solutions were annealed in K100 buffer. Each component was added in the order: 10 μl K100 buffer, 10 μl oligonucleotide and 5 μl of 10 μM ThT (dissolved in milli-Q water). The plate was shaken for 5 min and was incubated for 10 min at room temperature. Fluorescence intensity was collected at 490 nm after excitation at 420 nm in a TECAN M1000 pro plate reader.
- **NMM (*N*-methyl mesoporphyrin IX)** was used under the same condition as ThT, except that fluorescence intensity was collected at 610 nm after excitation at 380 nm in a TECAN M1000 pro plate reader.

G-Quadruplex ligands

The synthesis of G4 ligands tested against *S. mansoni* was previously described (35–37). Stock solutions were prepared at 10 mM in DMSO.

Antischistosomal activity of G-quadruplex ligands

Newly transformed schistosomula (NTS) drug assay. *S. mansoni* cercariae were collected from infected snails and mechanically transformed to newly transformed schistosomula (NTS). 30-40 NTS/well were incubated with the drugs for 72 h at 37°C, 5% CO_2 in a final well volume of 200–250 μl. Compounds were tested in triplicate and the highest concentration of DMSO (<1%) served as control. Evaluation was done by microscopic readout (Carl Zeiss, Germany, magnification 80x) as summarized in a previous publication (38).

Adult S. mansoni *drug assay.* Animal studies were carried out following Swiss national and cantonal regulations on animal welfare at the Swiss Tropical and Public Health Institute (Basel, Switzerland (Swiss TPH) (permission no. 2070). Female mice (NMRI; age 3 weeks; weight *ca*. 20–22 g) were purchased from Charles River, Germany. Mice were kept under environmentally controlled conditions (temperature ~25°C; humidity ~70%; 12 h light and 12 h dark cycle) with free access to water and rodent diet, and acclimatized for 1 week before infection.

Adult schistosomes were collected by mechanical picking from the hepatic portal system and mesenteric veins of mice 49-day post-infection with 100 *S. mansoni* cercariae. Worms were incubated with the compounds for 72 h. Wells with 1% DMSO served as negative controls. IC_{50} values were calculated using CalcuSyn Version 2.0 (Biosoft, Cambridge, UK). Phenotypes were evaluated under an inverted microscope and viability scores calculated (38).

RESULTS

We have selected 10 helminth organisms based on their impact on health or relevance as model species. From the accessible genomes, three of the most pathogenic species of Platyhelminths (*Schistosoma haematobium*, *Schistosoma japonicum* and *Schistosoma mansoni*) and three important Nematode species (*Strongyloides stercoralis*, *Trichuris trichiura*, and *Ascaris lumbricoides*) were selected. As reference organisms we have additionally selected three Platyhelminths (*Trichobilharzia regenti*, *Dibothriocephalus latus*, *Taenia asiatica*) and *Caenorhabditis elegans*, one of the best studied Nematodes. Both mitochondrial DNA (mtDNA) and nuclear genomes were analyzed by the G4Hunter algorithm for the presence of PQS.

PQS in mitochondrial DNA

At first, we analyzed the mitochondrial genomes of the 10 helminths listed above (Supplementary Table S1). mtDNA length varied from 13,608 to 15,003 bp, with a GC content between 23% (*S. stercoralis*) and 32% (*T. trichiura*). The results show that GC content is poorly correlated with the number of detected PQS: we found only one PQS sequence for the organism with the highest GC content (*T. trichiura*) while *S. stercoralis*, which has the mitochondrial genome with the lowest GC content in our dataset, has 6 PQS in its mtDNA. In total, we found 77 PQS with a G4Hunter score above 1.2, but none with a score above 1.4 (in other words, all motifs found are in the 1.2–1.4 interval). A 1.2 threshold is considered as a reasonable compromise to identify G4 prone motif (45); higher scores correspond to motifs capable of forming very stable quadruplexes. The majority of these sequences (40/77) were found in the mtDNA of a single species, *T. regenti.*

To analyze the PQS localization in mtDNA, we downloaded their annotations from NCBI and overlaid the PQS presence with these features (Supplementary Material 03). The organism with the highest number of PQS in mtDNA, *T. regenti,* has the majority of PQS in the repeat region of its mtDNA. However, PQS were also found in gene regions of this organism encoding cytochrome *c* oxidase subunit III, cytochrome *b*, NADH dehydrogenases subunits 4 and 2 and ATP synthase F0 subunit 6.

For *S. mansoni,* all (11 out of 11) PQS are located in the CDS of the genes coding for cytochrome c oxidase and NADH dehydrogenase subunit. This can be found in Supplementary Material 03 and Supplementary Table S1. The results demonstrate a nonrandom distribution of PQS in mtDNA of analyzed organisms according to PQS position and reveal that the prevalence of PQS is related mainly to production of NADH dehydrogenase subunits in *S. mansoni.* Supplementary Table S2 shows that *S. mansoni* is the only organism with an overlap of the PQS with the gene coding for the NADH dehydrogenase subunits 2, 5 and 4; the other two PQS are located inside the region coding for the cytochrome *c* oxidase subunits I and II.

To explore if the PQS are over-represented in regions coding NADH dehydrogenase subunits, we took feature tables containing annotations of known features found in mtDNA sequences and counted them. We then counted features with PQS. As mentioned before, the majority of the PQS is located inside the mitochondrial genes and coding sequences (CDS). PQS within the region coding for NADH dehydrogenase subunits are present in over 17% of all NADH dehydrogenase subunit coding regions and over 36% of all regions coding for cytochrome *c* oxidase subunits. In contrast, only 10% of rRNA and <2% of tRNA sequences contained PQS suggesting enrichment in NADH dehydrogenase sub-

unit and cytochrome *c* oxidase subunits. Data are available in Supplementary Material 03. PQS within the region coding for cytochrome c oxidase subunits are present in each analyzed organism. The position of this PQS may vary; it is found either in the region coding for the subunit I (*S. hematobium, S. japonicum and S. mansoni*) or subunit II (*S. mansoni*). A complete table can be found in Supplementary Material 03.

Predicted PQS in mtDNA of the three analyzed schistosoma species were compared looking for conserved motifs or sequences. We aligned the predicted PQS and generated their LOGO using the WEB LOGO tool (39) and the results are presented in Supplementary Material 04 for the most conserved sequences among the three *Schistosoma* species analyzed here: *S. haematobium, S. japonicum* and *S. mansoni*. A conservation of ~70% is found in these motifs. A similar level of conservation was seen among all six Platyhelminths. Evaluation of the PQS position shows that these most conserved sequences are in the COX1 gene.

PQS in nuclear DNA

Using standard values for G4Hunter (*i.e.* a window size of 25 nucleotides and threshold score of 1.2), we found over 1.3 million PQS among all 10 genomes. Overall, we do not advocate the use of a single threshold, but prefer to analyze data for various G4Hunter score windows. For this reason, we present these with different threshold windows (Table 1). We do have a good idea on false positive (FP) rate depending on threshold (<5% for scores >1.25; <2% for score >1.5 and close to 0 for scores >1.75; (45) and unpublished data; false positives are defined as sequences predicted to form a stable quadruplex, but cannot be confirmed experimentally). Our understanding on false negative (FN) rate as a function of threshold is less reliable, primarily because we did not explore as many sequences with relatively low G4Hunter scores (false negatives are sequences forming stable G4 which are missed by the algorithm). Our unpublished data with Dr Laurent Lacroix suggests that 1.2 is a reasonable threshold to maximize accuracy (*i.e.* minimize the fraction of FP + FN). For this reason, we chose this value as the minimal threshold considered. Higher thresholds (up to 2.0) tend to select sequences with high stability and propensity to form G4 structures.

Total PQS counts, percentage of GC and PQS frequencies characteristics for each organism are summarized in Table 2. The length of analyzed nuclear genomes varied from 42 Mbp (*S. stercoralis*) to 701 Mbp (*T. regenti*). The mean GC content was 34.9%, with a minimum of 22.1% for *S. stercoralis*, and a maximum of 42.2% for *T. trichiura*. The highest PQS frequencies were found in *D. latus* in which 473 thousand PQS were present in a 531 Mb genome (giving a PQS frequency of 0.89 PQS per 1000 nucleotides) and *T. asiatica* with 151,000 PQS in a 168 Mb genome, giving a similar PQS frequency of 0.90 PQS per kb (exact values provided in Table 2). In contrast, the parasite with the lowest PQS frequency was *S. stercoralis* with only 3,037 PQS for a 42.7 Mbp nuclear genome, giving a PQS frequency of 0.071 PQS per kb. We then compared the PQS frequency between moderately and highly infective parasites (40,41). Interestingly,

Table 1. Total number of PQS and their frequencies according to their G4Hunter score

Interval of G4Hunter score	Number of PQS in dataset	PQS frequency per 1000 bp
All		
1.2–1.4	1,281,682	0.396
1.4–1.6	29,381	0.010
1.6–1.8	5,631	0.002
1.8–2.0	2,158	0.001
2.0 or more	2,346	0.001
Platyhelminths		
1.2–1.4	1,051,587	0.440
1.4–1.6	16,715	0.007
1.6–1.8	976	≈0
1.8–2.0	42	≈0
2.0 or more	24	≈0
Nematodes		
1.2–1.4	230,095	0.329
1.4–1.6	12,666	0.014
1.6–1.8	4,655	0.004
1.8–2.0	2,116	0.002
2.0 or more	2,322	0.002

in these groups, the highest % of PQS was found in less infective Platyhelminths (with an average of 2.2 PQS per 1000 nt), followed by less infective Nematodes (1.6 PQS per kb (Table 2C). Detailed results are shown in Supplementary Material 02.

PQS frequency partially depends on GC content (a GC-poor genome is less likely to exhibit local G-rich motifs necessary for G4 formation). We present the global density in PQS (for all motifs with a G4Hunter score above 1.2) as a function of GC content for each individual organism in Figure 1A. A detailed analysis for each G4Hunter score interval is provided in Supplementary Material 06. The Spearman's rank correlation coefficient (r_s) was used to determine the association between PQS and GC content (Supplementary Material 05). PQS frequency is correlated with GC% ($r_s = 0.78$). This correlation is however far from perfect: for example, *T. trichiura* and *T. asiatica* have almost the same GC content, but their PQS frequencies differ considerably (0.263 for *T. trichiuria vs* 0.881 PQS per kb for *T. asiatica*). When considering all potential G4s (for a threshold ≥ 1.2), Platyhelminths look enriched with PQS (with a total of 1,069,344 PQS found) compared to Nematoda (with a total of 251,854 PQS found).

The results appear completely different if we restrict the analysis of PQS frequency to motifs with a G4Hunter score above 1.6 (Figure 1B)—in this case most species exhibit very low PQS frequencies (<0.002/kb), with the striking exception of *A. lumbricoides*, for which the density is at least 10-fold higher than in any other species.

In contrast with mitochondria (where no PQS with a score >1.4 was found), we still found a significant number of PQS with high G4Hunter scores (Table 1). However, and as for other species including bacteria (42) and archaea (43), the number of G4 motifs found drops significantly when higher thresholds are selected. This observation is valid for all genomes tested so far, including viruses, bacteria, archaea and eukaryotes. The majority of the PQS have a score in the range 1.2–1.4: most PQS sequences have a relatively

Table 2. G4Hunter analysis results for nuclear DNA. **(A)** Statistics for all tested organisms and clades. Seq (number of sequences), Median, Short and Long correspond to the median, min. and max. lengths in the dataset, GC% (GC content), PQS (number of PQS), Mean *f* (mean of PQS per kb), Min f (lowest frequency), Max f (highest frequency). **(B)** Statistics for individual organisms. **(C)** Statistics for organisms divided into highly and moderately infective categories

(A)

Domain	Seq	Median	Short	Long	PQS	Mean f	Min f	Max f	GC%
All	10	346,434,783	42,674,647	701,762,036	1,321,198	0.409	0.071	0.897	34.9
Group	**Seq**	**Median**	**Short**	**Long**	**PQS**	**Mean f**	**Min f**	**Max f**	**GC%**
Platyhelminths	6	406,161,092	168,679,183	701,762,036	1,069,344	0.448	0.190	0.897	35.6
Nematodes	4	87,891,397	42,674,647	316,975,410	251,854	0.350	0.071	0.563	34.1

(B)

Organism	Length	PQS	Mean f	GC%
Schistosoma haematobium	375,894,156	71,575	0.190	32.11
Schistosoma japonicum	402,743,189	84,683	0.210	31.23
Schistosoma mansoni	409,579,008	88,178	0.215	34.66
Trichobilharzia regenti	701,762,036	200,260	0.285	33.34
Dibothriocephalus latus	531,434,409	473,393	0.891	40.07
Taenia asiatica	168,679,183	151,255	0.897	42.06
Strongyloides stercoralis	42,674,647	20,471	0.071	22.11
Trichuris trichiura	75,496,394	3,037	0.271	42.22
Ascaris lumbricoides	316,975,410	178,583	0.563	36.48
Caenorhabditis elegans	100,286,401	49,763	0.496	35.44

(C)

Infectiousness	Seq	PQS	Mean PQS	Mean f	Mean PQS%	Mean PQS/GC%	Mean GC%
Platyhelminths High (Schistosomas spp)	3	244,436	81,478.67	0.205	0.665	358,253.80	32.66
Platyhelminths Low	3	824,908	274,969.33	0.691	2.229	713,947.83	38.49
Nematoda Highly (all except *C. elegans*)	3	202,091	67,363.67	0.302	1.060	183,916.89	33.60

low G4Hunter score. PQS in the next interval (1.4–1.6) are less frequent, and the drop continues for higher values (note the logarithmic axis on the Y-scale; Figure 2).

Compared to mtDNA, the nuclear DNAs are still poorly annotated in these species, with exception of *C. elegans*. Contrary to mtDNA, where PQS are located in the repeat region and gene regions, PQS are slightly enriched before and after gene regions, in the annotations for various RNA such as ncRNA, tRNA and precursor_RNA. The most significant enrichment for PQS were found in the regions before and after rRNA and prim_transcript (Supplementary Material 10: Elegans_Annotations-result.xls).

The comparison between phyla is interesting; *Homo sapiens*, archaea and bacteria are provided for comparison (43). Of note, PQS with a high G4Hunter score are very rare in Platyhelminths, while low-score PQS are extremely abundant. Platyhelminths behave like archaea and bacteria, with a stronger counter selection against very stable G4 than in *Homo sapiens* (Figure 2). Very stable G4 are therefore strongly counter selected in Platyhelminths as compared to humans: these G4 with high G4Hunter scores are extremely rare.

We then analyzed individual Nematode species in Figure 2B. We plotted data for each species (*Homo sapiens* and the group of three nematodes are again provided for comparison). As can be shown, *A. lumbricoides* is the organism that maintains the highest relative level of stable PQS at high thresholds, while the three other Nematoda exhibit a drop in relative number of PQS that is comparable (for *C. elegans*) or even sharper (for *S. stercoralis* and *T. trichiura*) than *Homo sapiens*.

A. lumbricoides is therefore unique among nematodes, and actually unique among all species we have studied so far: while its overall density in all G4 motifs is not remarkable (Figure 2), it can tolerate high-stability G4 structures, suggesting that these motifs are not counter-selected at all. The data is available in Supplementary Material 07.

PQS in Ascaris lumbricoides genome. The ability of *A. lumbricoides* to maintain the highest relative level of stable PQS at high thresholds prompted us to analyze its G4 motifs in more details. To this aim, we selected the 2,313 sequences with a G4Hunter score above 2.0 (Supplementary Material 08) and analyzed them.
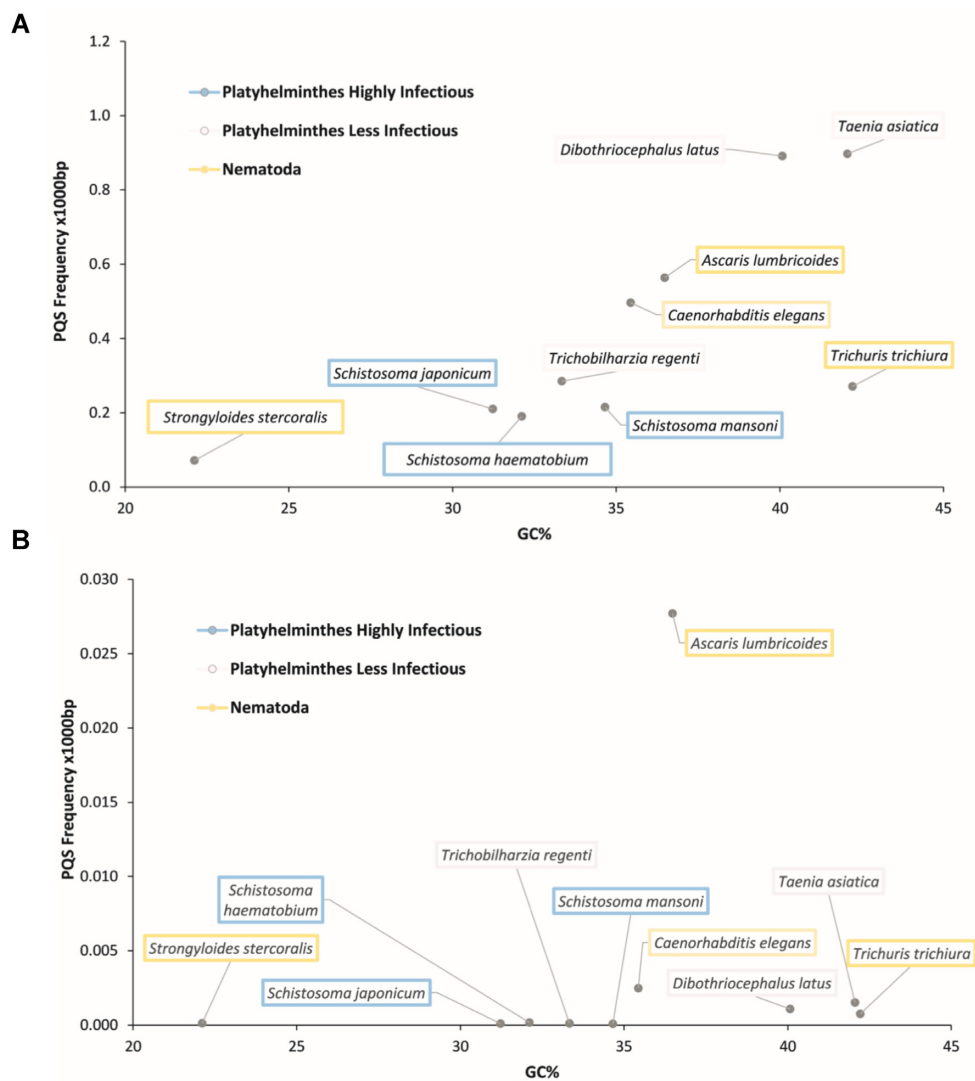
Figure 1. Relationship between GC% and PQS frequency per kb in genomic DNA (Color codes correspond to the groups (see Table 2, blue - Platyhelminths, yellow – Nematodes). Two different G4Hunter score thresholds were chosen for this analysis: (**A**) threshold ≥1.2. (**B**) threshold ≥1.6.

The most salient common feature of nearly all these *A. lumbricoides* motifs is that they are composed of poly-dG stretches, (or poly-dC when the score is negative) rather than other repetitive motifs (GGGT, GGGA, GGGGT, GGGGA, GGGGC, GGGGTT and GGGGAA would all give G4Hunter scores above 2.0). Out of 2,313 *A. lumbricoides* sequences with a G4Hunter score above 2.0 (or <2.0), 2311 (99.9%) contain at least one run of at least 10 C/10 G. The length distribution of these runs is presented in Figure 3.

We then checked if a similar behavior was found for nematodes closely related to *A. lumbricoides* (40) (Supplementary Table S4). *T. canis* contained the highest number of PQS: 242,923 with a frequency of 0.77 PQS/kb and a GC content of 37.6% in the total genome. *T. canis* is followed by *A. suum*, with 182,227 PQS, with a frequency of 0.61 PQS/kb and a GC% of 37.7% (Supplementary Table S5). The majority of the PQS was found in the G4Hunter threshold 1.2–1.4 (476,476 PQS in total) with a frequency of 0.259 PQS/kb inside this range.

Programmed DNA elimination is a feature of nematodes such as *A. lumbricoides* and *A. suum*. This developmentally regulated process leads to the reproducible loss of specific genomic sequences in somatic cells, leaving the germline genome intact. In-depth analyses of DNA elimination in *A. lumbricoides* and *T. canis* have been performed recently (40). We analyzed G4 propensity in the breakpoint regions of *A. suum*, and found that 39 out of 40 regions contained at least one PQS nearby (within 3 kb), suggesting a possible role of G-quadruplexes in this process. When it comes to the total frequency comparison, the majority of the PQS found in *A. suum* are present before or inside the chromosomal break regions.

To test whether the PQS occurrence in chromosomal breakpoint regions is significantly different than in randomly shuffled sequences, we generated 40 random sequences with the same length and nucleotide content as original *A. suum* breakpoints. We found that 38 out of 40 randomly generated regions contained at least one PQS (compared to the original breakpoint regions where PQS
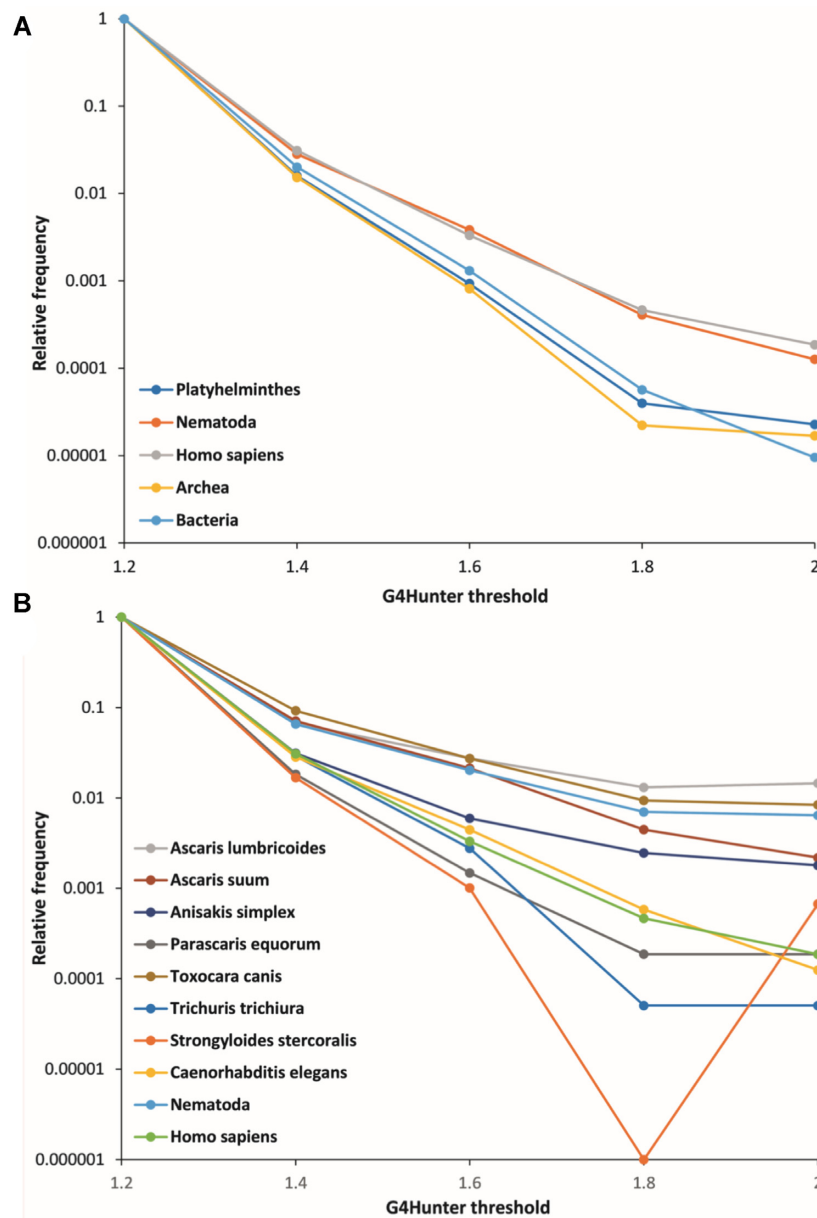
Figure 2. Distribution of G4 prone motifs per G4Hunter score: (**A**) in different phyla and (**B**) in selected species. *A. lumbricoides* exhibits an odd behavior (red curve) with a very high relative number of PQS with G4Hunter scores above 1.4. Data used to generate this panel can be found in Supplementary Table S3.

occurred in 39 out of 40 regions). However, the mean PQS count decreased from 9.25 PQS per breakpoint to 5.6 PQS. In addition, there was a highly significant difference between PQS counts ($P < 0.005$) in breakpoints and shuffled sequences. We performed a similar analysis with various G4Hunter thresholds (1.2–1.6). The higher the threshold, the higher the difference (and significance) between *A. suum* breakpoints and shuffled sequences. For a threshold of 1.6, 35 breakpoint sequences contain a least one PQS, while only 11 shuffled sequences contain one PQS (Figure 4; data provided in Sup Material 09).

To check if this trend was valid in other helminths, we performed additional analyzes of the breakpoints in two related species, *Toxocara canis* and *Parascaris equorum*,

thanks to the data collected by Wang *et al*. (44). We found a clear overrepresentation of G4 motifs in breakpoints as compared to shuffled sequences, and this result was valid at all threshold considered (see Supplementary Figure S7) as found in *A. suum*.

Experimental evidence for G4 formation *in vitro*

We identified a number of potential G4-forming motifs in helminth genomes using bioinformatics approaches. While G4Hunter's accuracy is reasonable, and actually excellent for high-scoring motifs (45), we found essential to confirm experimentally that some of these sequences are actually forming G4s, at least *in vitro*. To do so, we chose four se-
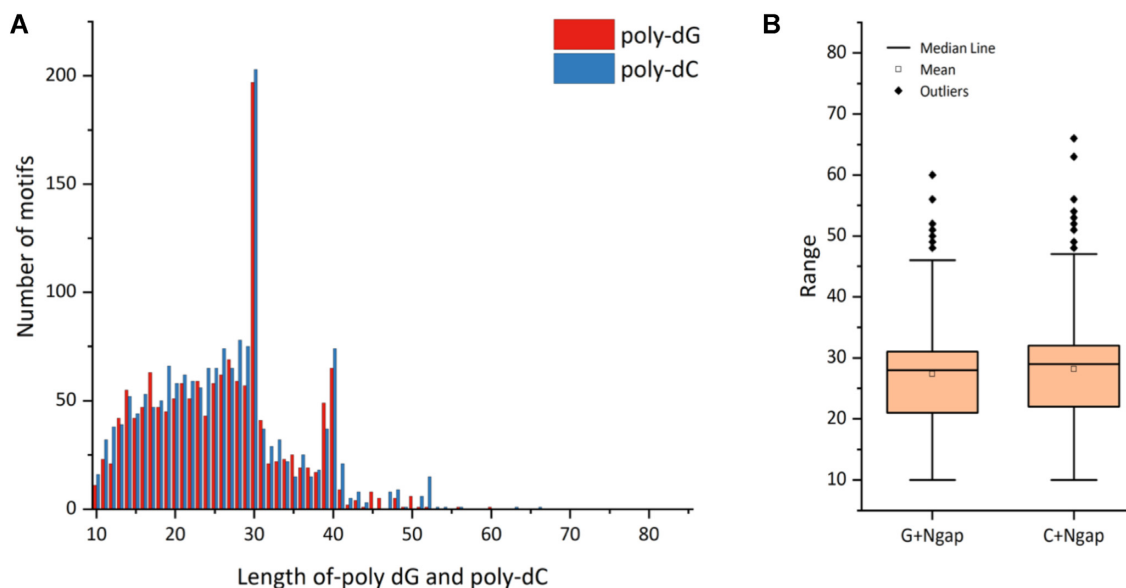
Figure 3. Distribution of pure homopolymeric G/C runs in *Ascaris lumbricoides* PQS motifs with a G4Hunter score above 2.0. (**A**) Length distribution for runs of 10 or more G (red) or C (blue). (**B**) Size distribution of G (left) and C (right) runs. Note that the genomic data include a number of N nucleotides corresponding to sequencing errors, adjacent to, or interrupting poly G/poly C runs. This is not very surprising, as reading error-prone sequences like these ones is difficult: this problem tends to underestimate the real length of poly dG stretches.

quences from *S. stercoralis*, three from *T. trichiura* as well as 20 sequences from *A. lumbricoides*. Results for *S. stercoralis* and *T. trichiura* are presented in Supplementary Information (Supplementary Table S6; Supplementary Figure S1-S2).

For the biophysical characterization of *A. lumbricoides* motifs, one should note that poly-dG runs have already been characterized in previous articles (46). For this reason, we focused our efforts on *A. lumbricoides* motifs that do not correspond to pure homopolymeric runs of guanines. These 20 sequences, which are 22–39 nucleotide-long, have G4Hunter scores between 1.46 and 2.48. G4 formation is considered extremely likely (>98.5%) for sequences with a score above 1.5 (45) but we wanted to provide compelling proof of G4 formation.

To demonstrate G4 formation, we used a combination of techniques, starting with FRET-MC, a method we very recently introduced (32) (Figure 5A). FRET-MC allows to test multiple sequences in parallel. A negative control (a sequence that does not form a G4 but a duplex, ds26) and two positive controls (sequences known to adopt G4 structures, Pu24T and c-myc) were used for comparison. The FRET-MC method measures the ability of a sequence to compete for binding to a well-known G4 ligand, PhenDC3. This compound is highly selective for G4s: efficient competitors are able to act as decoys for this G4 ligand, leading to a strong decrease in ΔT_m of a fluorescent G4-forming oligonucleotide (29); when added in large excess, these specific competitors can lead to negligible ΔT_m values. As can be seen in this panel, 19 out of 20 sequences considered here acted as efficient competitors (AL1 was found to be less efficient), arguing for G4 formation for most motifs.

Concluding on G4 formation based on a single technique is not recommended (45). Therefore, we used an independent approach, and investigated if the fluorescence emission

of G4 light-up probes (47), such as Thioflavin T (34) and NMM (48) was increased in the presence of *A. lumbricoides* motifs. Two negative and four positive controls were tested for comparison (Figure 5B, C). As shown in these panels, most sequences (including AL1 with NMM) induced significant increases in fluorescence emission to levels comparable or higher than the positive controls tested.

Finally, we performed additional spectroscopic experiments. G4s give specific signatures in isothermal and thermal difference spectra (IDS and TDS, respectively). The principle of these experiments is to compare the absorbance properties of the same oligonucleotide, in folded and unfolded state. The arithmetic difference between these two spectra gives a difference spectrum. Unfolding can be achieved by heating (for TDS) or by omitting stabilizing cations (for IDS). G4s exhibit a negative peak around 295 nm and a positive peak around 273 nm for both IDS and TDS (33). IDS and TDS of *Ascaris* motifs are shown in Supplementary Figure S3. Circular dichroism spectra of all AL sequences are presented in Supplementary Figure S4; interestingly, most spectra were indicative of parallel G4 formation (either exclusively or predominantly), except for AL14.

Altogether, these experiments confirmed predominant G4 formation for 19 out of the 20 *Ascaris* motifs tested. Of note, the motif for which a G4 may not be the dominant species and/or is of marginal stability (AL1) is the one with the lowest G4Hunter score (1.4).

Activity of G-quadruplex ligands against *Schistosoma mansoni*

We first verified that the *S. mansoni* G4 sequences recently described in (29) formed G4s, using the same techniques as the ones described for *A. lumbricoides*. As shown in Sup-
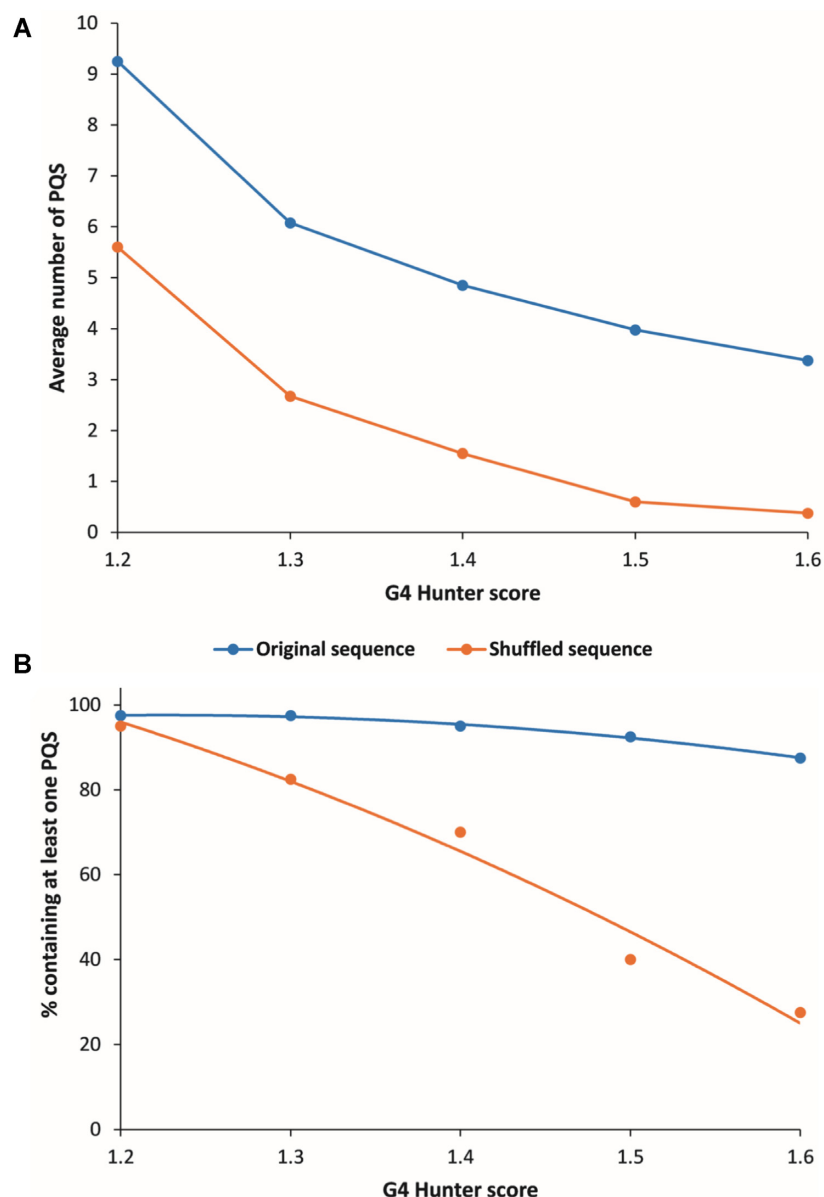
Figure 4. (**A**) Average number of PQS found in each breakpoint region, using different G4Hunter thresholds (x-avis) (**B**) fraction of breakpoint regions containing at least one quadruplex motif with a G4Hunter score above a certain threshold.

plementary Figure S5, this is indeed the case for the five positive sequences described in this article. Interestingly, for the sixth one (*smp-196840*), while the CD spectra reported by Craven *et al.* could not be associated with any known G4 CD profile, we were able to conclude that this motif was also able to form a stable G4, as shown by a combination of three independent methods (FRET-MC and two fluorescent light-up probes). The formation of a G4 by this motif is hardly surprising given its high G4Hunter score (2.0); *smp-196840* sequence is d-GGGAGGGGGGAGAGA GAGAGGGGGGAGGTAAAGGG). Overall, we conclude that all six sequences investigated form stable G4s.

We next investigated whether G4 ligands (i.e. small compounds which selectively recognize this unusual nucleic acid structure) recently synthesized (Figure 6A) (35–36) would

bind to the *S. mansoni* G4s described in (29). Six compounds were chosen, with variable levels of stabilization of G4 structures. We performed a FRET-melting assay, in which we measured the melting temperature of a dual-labeled fluorescent G4-forming oligonucleotide (F21T, corresponding to the human telomeric motif, but also to *S. mansoni* telomeres). The tested compounds have variable affinities for the telomeric motif, with ΔT_m of +0 to +14°C. To verify that the active compounds were also able to recognize other *S. mansoni* quadruplexes, we added these sequences as unlabeled competitors. As shown in Figure 6B, the addition of some, but not all of these oligonucleotides led to a decrease in the stabilization induced by the G4 ligand considered here. This indicates that motifs such as smp163240 or smp319480 were able to act as efficient 'de-
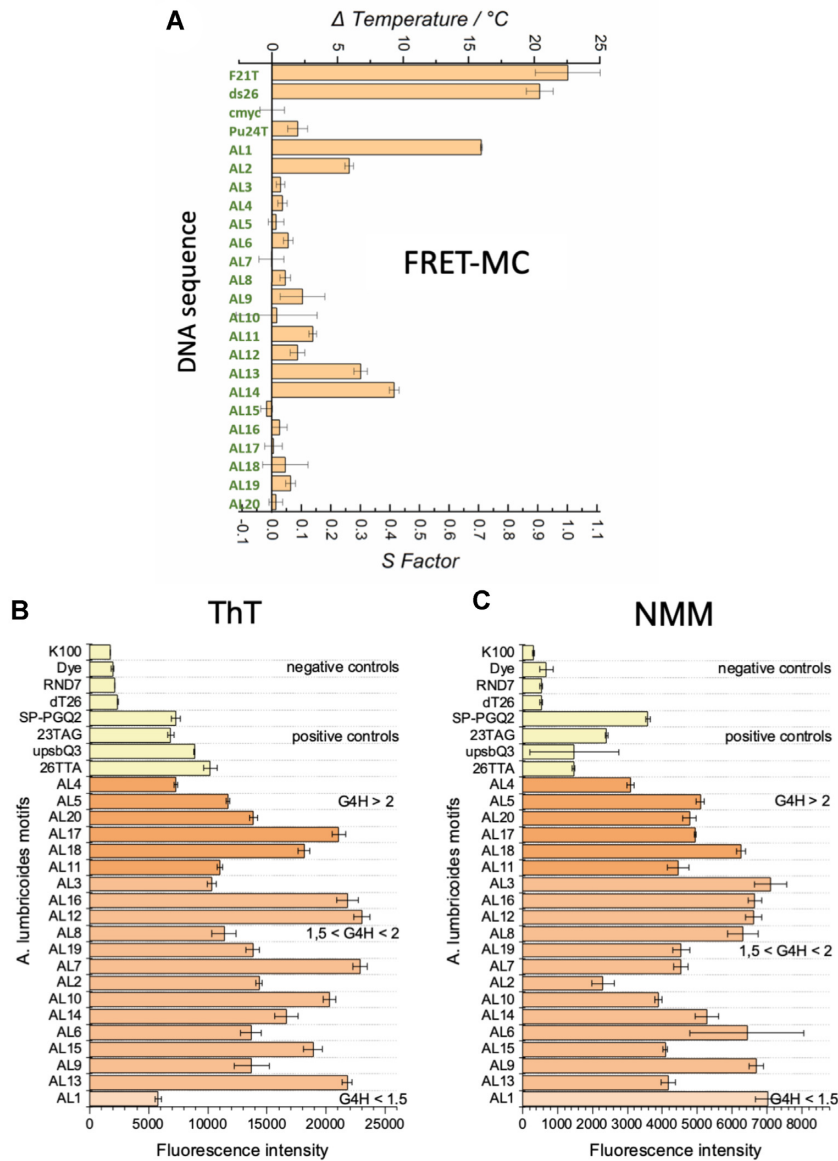
Figure 5. Evidence for G-quadruplex formation with 20 *Ascaris lumbricoides* sequences. (**A**) FRET MC results. (**B**) Thioflavin T fluorescence emission and (**C**) NMM fluorescence emission. Both Thioflavin T and NMM are light-up probes for which fluorescence intensity increases in the presence of G4-forming sequences.

coys' for the G4 ligands (even more efficient than the c-myc quadruplex used as a positive control), confirming that these molecules have an affinity for at least some of the G4 motifs (telomeric and non-telomeric) found in *S. mansoni* genome. In contrast, the ds26 negative control (double-stranded oligonucleotide) and two *S. mansoni* quadruplexes had little or no effect (Figure 6B), suggesting that JG1352 has little or no affinity for these structures. The FRET melting assay presented in Figure 7A illustrates that the best ligands (*e.g.* JG1352) stabilize all five G-quadruplexes tested in this assay to a variable extent, but do not stabilize the hairpin double-stranded control (FdxT; $\Delta T_\mathrm{m} \approx 0$; sequences provided in Supplementary Table S7).

The next step was to determine if these G4 ligands would have an antiparasitic activity. We tested the biological activ-

ity of these six compounds against larval and adult *S. mansoni*. JG1057 and JG1352 showed high activity at 100 μM and 10 μM and moderate activity at 1 μM against the larval stages. Both compounds revealed also high activity at 10 and 1 μM against adult *S. mansoni*. JG966 revealed a lower activity in particular against adult *S. mansoni*, affecting adult *S. mansoni* only at the highest concentration of 10 μM (Table 3). IC_{50} values calculated for JG1057 and JG1352 are in the range of praziquantel (a drug currently used to treat parasitic worm infections). Interestingly, the three inactive or weakly active compounds (ligands exhibiting low ΔT_m values on G-quadruplexes) (Figure 7A) were also significantly less active towards the parasite (Figure 7B) than two of the three best ligands. This effect was in particular against adult *S. mansoni* (both at 1 and 10 μM
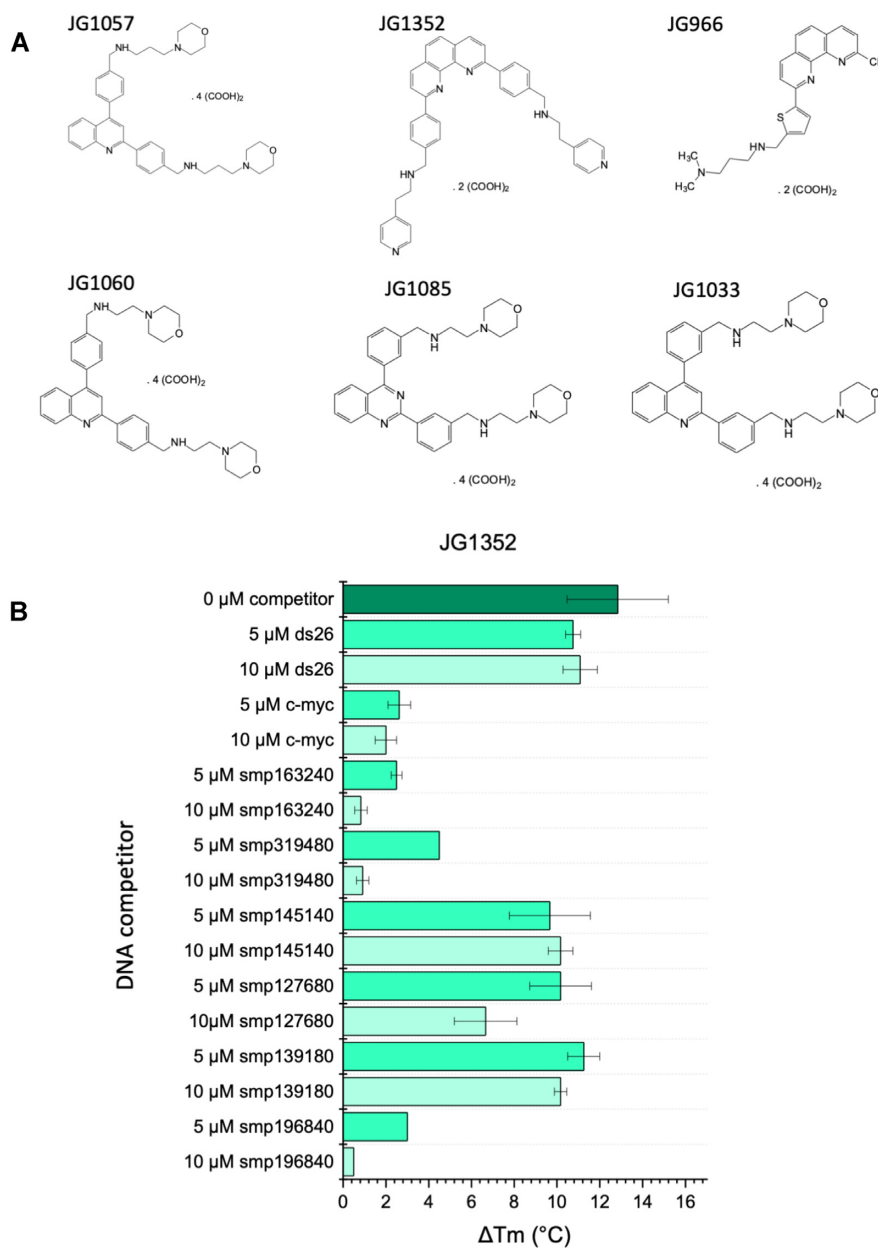
Figure 6. (A) Formula of the G4 ligands tested against *S. mansoni*. The three most active compounds (as determined by high ΔT_{m} values, see below) are shown on top; less active ligands are shown below. (B) Evidence that JG1352 binds to *S. mansoni* G-quadruplexes: FRET-melting data.

compound concentrations), suggesting that part of the antiparasitic effect of these compounds was mediated by a G4-related mechanism.

DISCUSSION

Besides the classical B-DNA double-helix structure, genomic DNA may adopt a variety of non-canonical structures which may play important roles (49). Repetitive sequences may form G4s (50) or i-DNA, inverted repeats can adopt cruciform structures (51), while CAG/CTG triplet repeats form unusual duplexes (52) and homopurine-homopyrimidine repeats adopt triplex structures (53). Many of these structures are involved in human patholo-

gies, such as neurological disorders or cancers (53–55). Recently their existence have been confirmed in living cells by several methods including secondary structures specific antibodies, synthetic compounds, and structure-sensitive sequencing (55–57). One of the most important local DNA structures seems to be G4, which has a better thermal stability compare to the B-DNA (58,59). Growing full-genome sequencing data provide an excellent source of information for detailed G4 prediction in various organisms. G4 has been shown to exist in archaea (43), bacteria (42,60,61) and eukaryote domains (19) as well in various viruses (62) where PQS propensity correlate with their host (63): viruses causing acute type of infections (including SARS-CoV2 genome) seem to be depleted for PQS (64,65). G4 have been
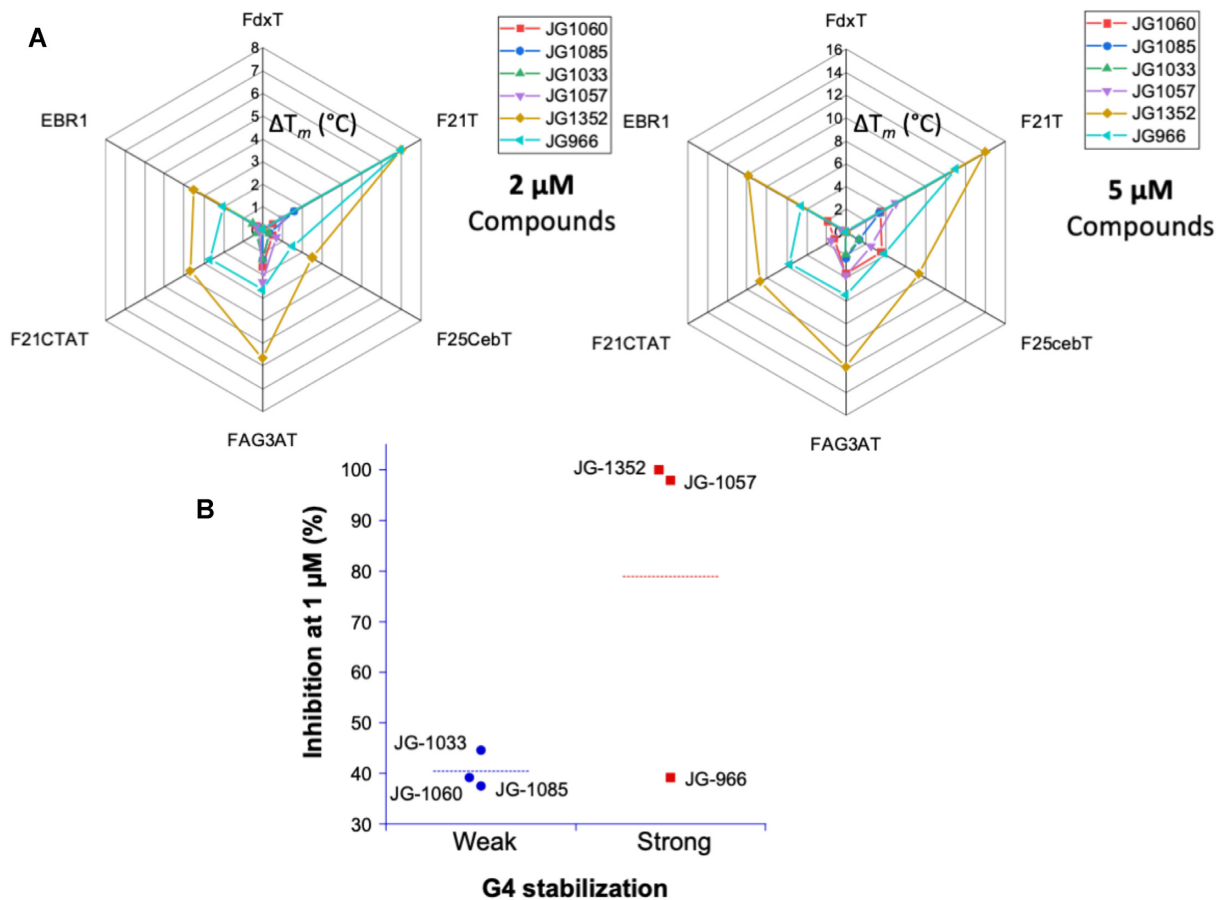
Figure 7. (**A**) ΔT_m values for the six compounds tested on six different nucleic acid structures, one duplex (FdxT) and five quadruplexes. Values are provided at two ligand concentrations, 2 (left) and 5 (right) μM. Sequences of the fluorescent oligonucleotides are provided in Supplementary Table S7. (**B**) Relation between activity against *S. mansoni* determined at 1 μM and G4 stabilization. Compounds were binned into two classes, 'weak' or 'strong', depending on ΔT_m on the human telomeric quadruplex (F21T; $\Delta T_m < 3.5°$C and $>5°$C, respectively).

Table 3. Antischistosomal activity of three G4 ligands tested at 3 different concentrations. Values correspond to growth inhibition (%). Compounds were tested in triplicate and *S. mansoni* incubated in the presence of the highest concentration of DMSO (<1%) served as control. The DMSO concentrations used in the assay (<1%) are routinely used in the lab, well tolerated and do not affect the parasite. Standard deviation is indicated within parentheses

Compound	Effect on newly transformed schistosomula (%) (SD)				Effect on adult *S.mansoni* (%) (SD)			
	100 μM	10 μM	1 μM	IC$_{50}$ value (μM)	10 μM	1 μM	0.1 μM	IC$_{50}$ value (μM)
JG 1057	100.0 (0)	92.5 (0.8)	46.7 (6.7)	1.49	98.2 (1.9)	97.9 (2.1)	29.6 (0)	0.11
JG 1352	100.0 (0)	100.0 (0)	56.7 (3.3)	0.4	100.0 (0)	100.0 (0)	31.5 (1.9)	0.07
JG 966	97.9 (2.1)	78.8 (7.9)	50.0 (10.0)	1.28	40.6 (0)	39.2 (0)	**ND**	>10
JG 1085	**ND**	100 (0)	32.7 (1.9)	1.18	48.2 (3.7)	37.4 (1.8)	**ND**	>10
JG 1060	**ND**	100 (0)	27.0 (3.8)	1.25	34.5 (3.6)	39.2 (3.6)	**ND**	>10
JG 1033	**ND**	100 (0)	25.0 (1.9)	1.28	48.2 (3.7)	44.6 (1.8)	**ND**	>10
Praziquantel[a]				2.2				0.1

[a]Reference for the IC$_{50}$ values of praziquantel taken from (80).

suggested to be valuable druggable targets for the development of a therapy against various pathogens (60,66). The recently published genome sequences of helminth species allowed us to perform comparative analyses of PQS in their genomes including broadly spread pathogenic species (4). We provided compelling evidence that many of these sequences form G4 structures *in vitro*. This does not necessarily imply that all of them adopt a G4 fold *in vivo*, as PQS were tested in the absence of flanking sequences, their reverse complement sequence, and proteins that could promote or inhibit folding. Nevertheless, results previously published on *S. mansoni* with the BG4 antibody attest that at least some of the candidate motifs adopt a quadruplex fold *in vivo* (29).

Comparison of PQS in nuclear and mtDNA

Far less potential G4 motifs are found in mitochondrial DNA. This is due to the much smaller size of the mitochondrial genome, even when considering the number of PQS

per kb, a lower density is still found on mtDNA, and this observation is true both for platyhelminths and nematodes (Supplementary Figure S6). Overall, and in stark contrast to humans, where G4 are abundant in mtDNA, G4 seem relatively rare in helminth mtDNA. Differences in GC content partially account for these differences. Nuclear DNA and mtDNA of *S. mansoni*, *S. japonicum*, *S. haematobium* and *T. regenti* have relatively similar GC content, while for other helminths (*D. latus*, *T. asiatica*, *T. trichiura*, *S. stercoralis*, *A. lumbricoides*, *C. elegans*), the GC content of nuclear DNA is higher by 8% or more over mtDNA. Surprisingly, only one parasite (*T. regenti*) has a higher mitochondrial over nuclear GC content.

This relative paucity in mtDNA prevents an in-depth comparison of mt G4 density between helminths. Nevertheless, some motifs appear conserved between species and are located in gene regions with potential to regulate their expression. On the other hand, numerous PQS are found in the genomes of all ten helminths considered here. Unfortunately, due to lack of annotation in the most species, the only species for which a detailed analysis of the reference genome is possible is *C. elegans* (67,68). These analyses of PQS localization would be interesting to evidence putative G4 regulatory function in translation, which have already been demonstrated in other organisms (69–71).

Telomeric motifs

While Platyhelminths and humans share the same telomeric motif (TTAGGG)$_n$ (72), Nematodes have a slightly different telomeric repeat (TTAGGC)$_n$ (73). This one-nucleotide difference has a strong impact on G4 formation. The human telomeric DNA motif (TTAGGG)$_n$ with a G4Hunter score of 1.5 is able to form a stable G4 both *in vitro* (74) and *in vivo* (75,76). Interestingly, both *S. haematobium* and *S. mansoni* have telomeric-like sequences integrated at non-telomeric sites (72). In contrast, the Nematode telomeric motif (TTAGGC)$_n$ has a significantly lower G4Hunter score (0.5). Despite this low score, such motif may form an antiparallel G4 ($T_m \approx 40°C$; Marquevielle, submitted), possibly in competition with a fundamentally different secondary structure called a foldback (77).

Programmed DNA elimination in nematodes

DNA elimination occurs in a number of species, including nematodes. It mostly corresponds to repetitive sequences and germline-specific genes. Previous analyses suggest that DNA elimination in nematodes silences germline-expressed genes (44). Their results suggested a sequence-independent mechanism for DNA breakage. Interestingly, we found that there is a clear overrepresentation of G4 motifs in breakpoints as compared to shuffled sequences, and this result was valid at all threshold considered (see Figure 4 and Supplementary Figure S7; Supplementary Material 9) and for three nematode species: *Ascaris suum*, *Toxocara canis* and *Parascaris equorum*. This conservation suggests that G-quadruplex formation may be involved in this programmed elimination, perhaps by recruiting a DNA cleaving complex (44). Telomere healing (that would lead to the insertion of a few telomeric motifs) would not explain this bias given the low G4Hunter score (see § above) of nematode telomeres.

Ascaris lumbricoides, a unique organism

Using G4Hunter score as a proxy for G4 stability, there is a strong counterselection against stable G4 in all Platyhelminths, in a manner similar to what is found in archaea and bacteria. G4 motifs with G4Hunter scores above 1.8 or 2.0 tend to be very rare, in comparison with the total number of motifs, as illustrated in Figure 2. This may suggest that Platyhelminths are unable to cope with very stable G4s, which cause problems during replication or transcription. On the other hand, nematodes follow a profile comparable to humans, where selection against stable G4 is not as strong as in prokaryotes.

Poly G runs frequency is worth discussing in the light of the analysis performed by Puig-Lombardi *et al.* (78). They analyzed the frequency of (GGGN)$_3$GGG motifs (15-nt sequence), corresponding to four runs of three guanines plus one extra nucleotide N. When N = G, this corresponds to 'pure' G_{15} runs. This analysis was performed on over hundreds of genome assemblies. The authors found that the (GGGA)$_3$GGG motif was largely predominant in placental mammals, including humans, while an excess of (GGGG)$_3$GGG (pure poly G runs) was found in amphibians, fish, plants, invertebrates and nematodes, including *C. elegans* (see Fig. S15 of reference (78)). The authors wrote that 'the extreme prevalence (98%) of the G-runs versus the other GGGX motifs in the *C. elegans* genome and the finding that these sequences are eliminated by complete deletion during development and in animals deficient for the dog-1 helicase suggest that different molecular mechanisms can play a role in handling the equilibrium between the maintenance and the inactivation of short-loop G4-L1 motifs'.

What is unique about *A. lumbricoides* is therefore not the presence of pure poly G runs as compared to other repeats, but the overall abundancy of long PQS with high G4Hunter scores. *A. lumbricoides* and related species such as *A. suum* have an exceptional density of high stability G4s (with a high G4Hunter score); far more frequent than any other species. These results suggest that *A. lumbricoides* must have evolved very efficient mechanisms to cope with these stable secondary structures. Therefore, we checked if its genome contained putative helicases susceptible to unfold G4s. Dog-1 has been found to be involved in the genomic stability of poly dG stretches in *C. elegans*. For this reason, we looked for orthologs of the Dog-1 helicase domain (aa 80–440) in *A. lumbricoides*. Interestingly, four genes had homology, suggesting putative helicases involved in G4 resolution/unfolding (see Supplementary Table S8). Of note, one of them is an ortholog of a human ATP-dependent DNA helicase (also called DDX11, CHLR1 or KRG2) reported to act on G4 substrates (79). Further studies could evaluate the impact of G4 ligands on *Ascaris* spp.

CONCLUSION

Helminth genomes contain multiple PQS sequences. The comparison of various nematodes and Platyhelminths revealed interesting and marked differences between helminths. Experimental confirmation of G4 formation *in vitro* was obtained for four different species. Two of the G4 ligands able to bind to *Schistosoma mansoni* G4 motifs exhibited potent antiparasitic activity both against larval

and adult forms of this parasite, opening new perspectives for the use of G4 ligands to fight neglected tropical diseases. These results therefore open new perspectives for the development of novel therapeutic strategies against these widespread helminths.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Hotez,P.J., Bundy,D.A.P., Beegle,K., Brooker,S., Drake,L., de Silva,N., Montresor,A., Engels,D., Jukes,M., Chitsulo,L. *et al.* (2006) Helminth Infections: soil-transmitted helminth infections and schistosomiasis. In: Jamison,D.T., Breman,J.G., Measham,A.R., Alleyne,G., Claeson,M., Evans,D.B., Jha,P., Mills,A. and Musgrove,P. (eds). *Disease Control Priorities in Developing Countries.* 2nd edn. Washington (DC).
2. Bethony,J., Brooker,S., Albonico,M., Geiger,S.M., Loukas,A., Diemert,D. and Hotez,P.J. (2006) Soil-transmitted helminth infections: ascariasis, trichuriasis and hookworm. *The Lancet.*, **367**, 1521–1532.
3. Verjee,MA. (2019) Schistosomiasis: still a cause of significant morbidity and mortality. *Res. Rep. Trop. Med.*, **10**, 153–163.
4. Anderson,T.J.C. and Duraisingh,M.T. (2020) Transformative tools for parasitic flatworms. *Science*, **369**, 1562–1564.
5. Christian,P., Khatry,S.K. and West,K.P. (2004). Antenatal anthelmintic treatment, birthweight, and infant survival in rural Nepal. *Lancet*, **364**, 981–983.
6. Charlier,J., De Waele,V., Ducheyne,E., van der Voort,M., Vande Velde,F. and Claerebout,E. (2016) Decision making on helminths in cattle: diagnostics, economics and human behaviour. *Irish Vet. J.*, **69**, 14.
7. de Silva,N.R., Brooker,S., Hotez,P.J., Montresor,A., Engels,D. and Savioli,L. (2003) Soil-transmitted helminth infections: updating the global picture. *Trends Parasitol.*, **19**, 547–551.
8. Nutman,T.B. (2017) Human infection with *Strongyloides stercoralis* and other related *Strongyloides* species. *Parasitology*, **144**, 263–273.
9. Colley,D.G., Bustinduy,A.L., Secor,W.E. and King,C.H. (2014) Human schistosomiasis. *Lancet.*, **383**, 2253–2264.
10. Thétiot-Laurent,S.A.-L., Boissier,J., Robert,A. and Meunier,B. (2013) Schistosomiasis chemotherapy. *Angew. Chem. Int. Ed.*, **52**, 7936–7956.
11. Furtado,L.F.V., de Paiva Bello,A.C.P. and Rabelo,É.M.L. (2016) Benzimidazole resistance in helminths: from problem to diagnosis. *Acta Tropica*, **162**, 95–102.
12. Bergquist,R., Zhou,X.-N., Rollinson,D., Reinhard-Rupp,J. and Klohe,K. (2017) Elimination of schistosomiasis: the tools required. *Infect. Dis. Poverty*, **6**, 158.
13. Berriman,M., Haas,B.J., LoVerde,P.T., Wilson,R.A., Dillon,G.P., Cerqueira,G.C., Mashiyama,S.T., Al-Lazikani,B., Andrade,L.F., Ashton,P.D. *et al.* (2009) The genome of the blood fluke *Schistosoma mansoni*. *Nature*, **460**, 352–358.
14. Zhou,Y., Zheng,H., Chen,Y., Zhang,L., Wang,K., Guo,J., Huang,Z., Zhang,B., Huang,W., Jin,K. *et al.* (2009) The *Schistosoma japonicum* genome reveals features of host–parasite interplay. *Nature*, **460**, 345–351.
15. Young,N.D., Jex,A.R., Li,B., Liu,S., Yang,L., Xiong,Z., Li,Y., Cantacessi,C., Hall,R.S., Xu,X. *et al.* (2012) Whole-genome sequence of *Schistosoma haematobium*. *Nat. Genet.*, **44**, 221–225.
16. Kolesnikova,S. and Curtis,E.A. (2019) Structure and function of multimeric G-quadruplexes. *Molecules*, **24**, 3074.
17. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
18. Lane,A.N., Chaires,J.B., Gray,R.D. and Trent,J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.
19. Yoshida,W., Saikyo,H., Nakabayashi,K., Yoshioka,H., Bay,D.H., Iida,K., Kawai,T., Hata,K., Ikebukuro,K., Nagasawa,K. *et al.* (2018) Identification of G-quadruplex clusters by high-throughput sequencing of whole-genome amplified products with a G-quadruplex ligand. *Sci. Rep.*, **8**, 3116.
20. Carvalho,J., Mergny,J.L., Salgado,G.F., Queiroz,J.A. and Cruz,C. (2020) G-quadruplex, friend or foe: the role of the G-quartet in anticancer strategies. *Trends Mol. Med.*, **26**, 848–861.
21. Balasubramanian,S., Hurley,L.H. and Neidle,S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–275.
22. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11593–11598.
23. Sun,Z.-Y., Wang,X.-N., Cheng,S.-Q., Su,X.-X. and Ou,T.-M. (2019) Developing novel G-quadruplex ligands: from interaction with nucleic acids to interfering with nucleic acid–protein interaction. *Molecules*, **24**, 396.
24. Zhao,J., Bacolla,A., Wang,G. and Vasquez,K.M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.*, **67**, 43–62.
25. Wong,H.M. and Huppert,J.L. (2009). Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. BioSys.*, **5**, 1713–1719.
26. Marsico,G., Chambers,V.S., Sahakyan,A.B., McCauley,P., Boutell,J.M., di Antonio,M. and Balasubramanian,S. (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.*, **47**, 3862–3874.
27. Hoffmann,R.F., Moshkin,Y.M., Mouton,S., Grzeschik,N.A., Kalicharan,R.D., Kuipers,J., Wolters,A.H.G., Nishida,K., Romashchenko,A.V., Postberg,J. *et al.* (2016) Guanine quadruplex structures localize to heterochromatin. *Nucleic Acids Res.*, **44**, 152–163.
28. Yu,Q.-Q., Gao,J.-J., Lang,X.-X., Li,H.-Y. and Wang,M.-Q. (2021) Microenvironment-sensitive fluorescent ligand binds ascaris telomere antiparallel G-quadruplex DNA with blue-shift and enhanced emission. *ChemBioChem.*, **22**, 1042–1048.
29. Craven,H.M., Bonsignore,R., Lenis,V., Santi,N., Berrar,D., Swain,M., Whiteland,H., Casini,A. and Hoffmann,K.F. (2021) Identifying and validating the presence of guanine-quadruplexes (G4) within the blood fluke parasite *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.*, **15**, e0008770.
30. Brázda,V., Kolomazník,J., Lýsek,J., Bartas,M., Fojta,M., Šťastný,J. and Mergny,J.L. (2019) G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics*, **35**, 3493–3495.
31. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
32. Luo,Y., Granzhan,A., Verga,D. and Mergny,J.-L. (2021) FRET-MC: a fluorescence melting competition assay for studying G4 structures in vitro. *Biopolymers*, **112**, e23415.
33. Mergny,J.-L., Li,J., Lacroix,L., Amrane,S. and Chaires,J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
34. Renaud de la Faverie,A., Guédin,A., Bedrat,A., Yatsunyk,L.A. and Mergny,J.-L. (2014) Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Res.*, **42**, e65.
35. Guillon,J., Denevault-Sabourin,C., Chevret,E., Brachet-Botineau,M., Milano,V., Guédin-Beaurepaire,A., Moreau,S., Ronga,L.,

Savrimoutou,S., Rubio,S. *et al.* (2021) Design, synthesis, and antiproliferative effect of 2,9-bis[4-(pyridinylalkylaminomethyl)phenyl]-1,10-phenanthroline derivatives on human leukemic cells by targeting G-quadruplex. *Archiv. Pharmazie*, **354**, e2000450.

36. Guillon,J., Cohen,A., Das,R.N., Boudot,C., Gueddouda,N.M., Moreau,S., Ronga,L., Savrimoutou,S., Basmaciyan,L., Tisnerat,C. *et al.* (2018) Design, synthesis, and antiprotozoal evaluation of new 2,9-bis[(substituted-aminomethyl)phenyl]-1,10-phenanthroline derivatives. *Chem. Biol. Drug Des.*, **91**, 974–995.

37. Guillon,J., Cohen,A., Boudot,C., Valle,A., Milano,V., Das,R.N., Guédin,A., Moreau,S., Ronga,L., Savrimoutou,S. *et al.* (2020) Design, synthesis, and antiprotozoal evaluation of new 2,4-bis[(substituted-aminomethyl)phenyl]quinoline, 1,3-bis[(substituted-aminomethyl)phenyl]isoquinoline and 2,4-bis[(substituted-aminomethyl)phenyl]quinazoline derivatives. *J. Enz. Inhib. Med. Chem.*, **35**, 432–459.

38. Lombardo,F.C., Pasche,V., Panic,G., Endriss,Y. and Keiser,J. (2019) Life cycle maintenance and drug-sensitivity assays for early drug discovery in *Schistosoma mansoni*. *Nat. Protoc.*, **14**, 461–481.

39. Crooks,G.E., Hon,G., Chandonia,J.-M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.

40. Coghlan,A., Tyagi,R., Cotton,J.A., Holroyd,N., Rosa,B.A., Tsai,I.J., Laetsch,D.R., Beech,R.N., Day,T.A., Hallsworth-Pepin,K. *et al.* (2019) Comparative genomics of the major parasitic worms. *Nat. Gen.*, **51**, 163–174.

41. Jourdan,P.M., Lamberton,P.H.L., Fenwick,A. and Addiss,D.G. (2018) Soil-transmitted helminth infections. *Lancet*, **391**, 252–265.

42. Bartas,M., Čutová,M., Brázda,V., Kaura,P., Šťastný,J., Kolomazník,J., Coufal,J., Goswami,P., Červen,J. and Pečinka,P. (2019) The presence and localization of G-quadruplex forming sequences in the domain of bacteria. *Molecules*, **24**, 1711.

43. Brázda,V., Luo,Y., Bartas,M., Kaura,P., Porubiaková,O., Šťastný,J., Pečinka,P., Verga,D., Da Cunha,V., Takahashi,T.S. *et al.* (2020) G-Quadruplexes in the archaea domain. *Biomolecules*, **10**, 1349.

44. Wang,J., Gao,S., Mostovoy,Y., Kang,Y., Zagoskin,M., Sun,Y., Zhang,B., White,L.K., Easton,A., Nutman,T.B. *et al.* (2017) Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Res.*, **27**, 2001–2014.

45. Bedrat,A., Lacroix,L. and Mergny,J.-L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.*, **44**, 1746–1759.

46. Sengar,A., Heddi,B. and Phan,A.T. (2014) Formation of G-quadruplexes in poly-G sequences: structure of a propeller-type parallel-stranded G-quadruplex formed by a G15 stretch. *Biochemistry*, **53**, 7718–7723.

47. Largy,E., Granzhan,A., Hamon,F., Verga,D. and Teulade-Fichou,M.P. (2013) Visualizing the quadruplex: from fluorescent ligands to light-up probes. *Top. Curr. Chem.*, **330**, 111–177.

48. Sabharwal,N.C., Savikhin,V., Turek-Herman,J.R., Nicoludis,J.M., Szalai,V.A. and Yatsunyk,L.A. (2014) N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. *FEBS J.*, **281**, 1726–1737.

49. McKinney,J.A., Wang,G., Mukherjee,A., Christensen,L., Subramanian,S.H.S., Zhao,J. and Vasquez,K.M. (2020) Distinct DNA repair pathways cause genomic instability at alternative DNA structures. *Nat Commun.*, **11**, 236.

50. Spiegel,J., Adhikari,S. and Balasubramanian,S. (2020) The structure and function of DNA G-quadruplexes. *Trends Chem.*, **2**, 123–136.

51. Brázda,V., Laister,R.C., Jagelská,E.B. and Arrowsmith,C. (2011) Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol Biol.*, **12**, 33.

52. Petruska,J., Arnheim,N. and Goodman,MF. (1996) Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Res.*, **24**, 1992–1998.

53. Helma,R., Bažantová,P., Petr,M., Adámik,M., Renčiuk,D., Tichý,V., Pastuchová,A., Soldánová,Z., Pečinka,P., Bowater,R.P. *et al.* (2019) p53 binds preferentially to non-B DNA structures formed by the pyrimidine-rich strands of GAA·TTC trinucleotide repeats associated with Friedreich's ataxia. *Molecules*, **24**, 2078.

54. Cimino-Reale,G., Zaffaroni,N. and Folini,M. (2016) Emerging role of G-quadruplex DNA as target in anticancer therapy. *Curr. Pharm. Des.*, **22**, 6612–6624.

55. Hänsel-Hertsch,R., Simeone,A., Shea,A., Hui,W.W.I., Zyner,K.G., Marsico,G., Rueda,O.M., Bruna,A., Martin,A., Zhang,X. *et al.* (2020) Landscape of G-quadruplex DNA structural regions in breast cancer. *Nat. Genet.*, **52**, 878–883.

56. Poggi,L. and Richard,G-F. (2021) Alternative DNA structures in vivo: molecular evidence and remaining questions. *Microbiol. Mol. Biol. Rev.*, **85**, e00110-20.

57. Yang,C., Hu,R., Li,Q., Li,S., Xiang,J., Guo,X., Wang,S., Zen,Y. and Yang,G. (2018) Visualization of parallel G-quadruplexes in cells with a series of new developed Bis(4-aminobenzylidene)acetone derivatives. *ACS Omega*, **3**, 10487–10492.

58. Bochman,M.L., Paeschke,K. and Zakian,V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.*, **13**, 770–780.

59. Takahashi,S. and Sugimoto,N. (2013) Effect of pressure on thermal stability of G-quadruplex DNA and double-stranded DNA structures. *Molecules*, **18**, 13297–13319.

60. Yadav,P., Kim,N., Kumari,M., Verma,S., Sharma,T.K., Yadav,V. and Kumar,A. (2021) G-Quadruplex structures in bacteria - biological relevance and potential as antimicrobial target. *J. Bacteriol.*, **203**, e0057720.

61. Brázda,V., Fojta,M. and Bowater,R.P. (2020) Structures and stability of simple DNA repeats from bacteria. *Biochem J.*, **477**, 325–339.

62. Ruggiero,E. and Richter,S.N. (2020) Viral G-quadruplexes: new frontiers in virus pathogenesis and antiviral therapy. *Annu. Rep. Med. Chem.*, **54**, 101–131.

63. Bohálová,N., Cantara,A., Bartas,M., Kaura,P., Šťastný,J., Pečinka,P., Fojta,M. and Brázda,V. (2021) Tracing dsDNA virus-host coevolution through correlation of their G-quadruplex-forming sequences. *Int. J. Mol. Sci.*, **22**, 3433.

64. Bohálová,N., Cantara,A., Bartas,M., Kaura,P., Šťastný,J., Pečinka,P., Fojta,M., Mergny,J.L. and Brázda,V. (2021) Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection. *Biochimie*, **186**, 13–27.

65. Bartas,M., Brázda,V., Bohálová,N., Cantara,A., Volná,A., Stachurová,T., Malachová,K., Jagelská,E.B., Porubiaková,O., Červeň,J. *et al.* (2020) In-depth bioinformatic analyses of human SARS-CoV-2, SARS-CoV, MERS-CoV, and other nidovirales suggest important roles of noncanonical nucleic acid structures in their lifecycles. *Front. Microbiol.*, **11**, 1583.

66. Ruggiero,E. and Richter,S.N. (2020) Viral G-quadruplexes: new frontiers in virus pathogenesis and antiviral therapy. *Annu. Rep. Med. Chem.*, **54**, 101–131.

67. Li,R., Hsieh,C.-L., Young,A., Zhang,Z., Ren,X. and Zhao,Z. (2015) Illumina synthetic long read sequencing allows recovery of missing sequences even in the «Finished» *C. elegans* genome. *Sci Rep.*, **5**, 10814.

68. Kruisselbrink,E., Guryev,V., Brouwer,K., Pontier,D.B., Cuppen,E. and Tijsterman,M. (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective *C. elegans*. *Curr Biol.*, **18**, 900–905.

69. Vannutelli,A., Belhamiti,S., Garant,J.-M., Ouangraoua,A. and Perreault,J.-P. (2020) Where are G-quadruplexes located in the human transcriptome? *NAR Genom. Bioinform.*, **2**, lqaa035.

70. Maltby,C.J., Schofield,J.P.R., Houghton,S.D., O'Kelly,I., Vargas-Caballero,M., Deinhardt,K. and Coldwell,M.J. (2020) A 5' UTR GGN repeat controls localisation and translation of a potassium leak channel mRNA through G-quadruplex formation. *Nucleic Acids Res.*, **48**, 9822–9839.

71. Katsuda,Y., Sato,S.-I., Asano,L., Morimura,Y., Furuta,T., Sugiyama,H., Hagihara,M. and Uesugi,M. (2016) A small molecule that represses translation of G-quadruplex-containing mRNA. *J. Am. Chem. Soc.*, **138**, 9037–9040.

72. Hirai,H. (2014) Chromosomal differentiation of schistosomes: what is the message? *Front. Genet.*, **5**, 301.

73. Müller,F., Wicky,C., Spicher,A. and Tobler,H. (1991) New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in *Ascaris lumbricoides*. *Cell*, **67**, 815–822.

74. Phan,A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.*, **277**, 1107–1117.

75. Hänsel,R., Foldynová-Trantírková,S., Löhr,F., Buck,J., Bongartz,E., Bamberg,E., Schwalbe,H., Dötsch,V. and Trantirek,L. (2009) Evaluation of parameters critical for observing nucleic acids inside living *Xenopus laevis* oocytes by in-cell NMR spectroscopy. *J. Am. Chem. Soc.*, **131**, 15761–15768.

76. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.

77. Školáková,P., Foldynová-Trantírková,S., Bednářová,K., Fiala,R., Vorlíčková,M. and Trantírek,L. (2015) Unique *C. elegans* telomeric overhang structures reveal the evolutionarily conserved properties of telomeric DNA. *Nucleic Acids Res*. **43**, 4733–4745.

78. Puig Lombardi,E., Holmes,A., Verga,D., Teulade-Fichou,M.-P., Nicolas,A. and Londoño-Vallejo,A. (2019) Thermodynamically stable and genetically unstable G-quadruplexes are depleted in genomes across species. *Nucleic Acids Res.*, **47**, 6098–6113.

79. Ribeiro de Almeida,C., Dhir,S., Dhir,A., Moghaddam,A.E., Sattentau,Q., Meinhart,A. and Proudfoot,N.J. (2018) RNA helicase DDX1 converts RNA G-quadruplex structures into R-loops to promote IgH class switch recombination. *Mol. Cell.*, **70**, 650–662.

80. Ingram,K., Yaremenko,I.A., Krylov,I.B., Hofer,L., Terent'ev,A.O. and Keiser,J. (2012) Identification of antischistosomal leads by evaluating bridged 1,2,4,5-tetraoxanes, alphaperoxides, and tricyclic monoperoxides. *J. Med. Chem.*, **55**, 8700–8711.