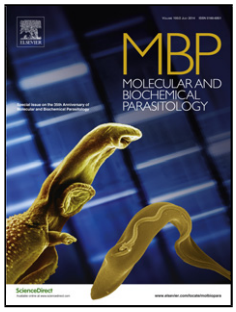


Title: Different SNPs in *Fasciola hepatica* P-glycoprotein from diverse Latin American populations are not associated with Triclabendazole resistance

Authors: María Victoria Solana, María Fernanda Domínguez, Silvana Scarcella, Santiago Radio, Pablo Smircich, Silvina Fernández, Hugo Solana, José F. Tort

PII: S0166-6851(18)30041-0
DOI: https://doi.org/10.1016/j.molbiopara.2018.07.005
Reference: MOLBIO 11138

To appear in: *Molecular & Biochemical Parasitology*

Received date: 20-2-2018
Revised date: 11-7-2018
Accepted date: 11-7-2018

Please cite this article as: Solana MV, Domínguez MF, Scarcella S, Radio S, Smircich P, Fernández S, Solana H, Tort JF, Different SNPs in *Fasciola hepatica* P-glycoprotein from diverse Latin American populations are not associated with Triclabendazole resistance, *Molecular and Biochemical Parasitology* (2018), https://doi.org/10.1016/j.molbiopara.2018.07.005

Different SNPs in *Fasciola hepatica* P-glycoprotein from diverse Latin American populations are not associated with Triclabendazole resistance

María Victoria Solana [a], María Fernanda Domínguez, [b], Silvana Scarcella [a], Santiago Radio [b, 1], Pablo Smircich [b, 1], Silvina Fernández [c], Hugo Solana [a#], José F. Tort [b #]

[a-] *Laboratorio de Biología Celular y Molecular, Centro de Investigación Veterinaria de Tandil (CIVETAN), CONICET, Facultad de Ciencias Veterinarias, UNCPBA, Tandil, Argentina.*

[b-] *Departamento de Genética, Facultad de Medicina, Universidad de la Republica, UDELAR, Montevideo, Uruguay.*

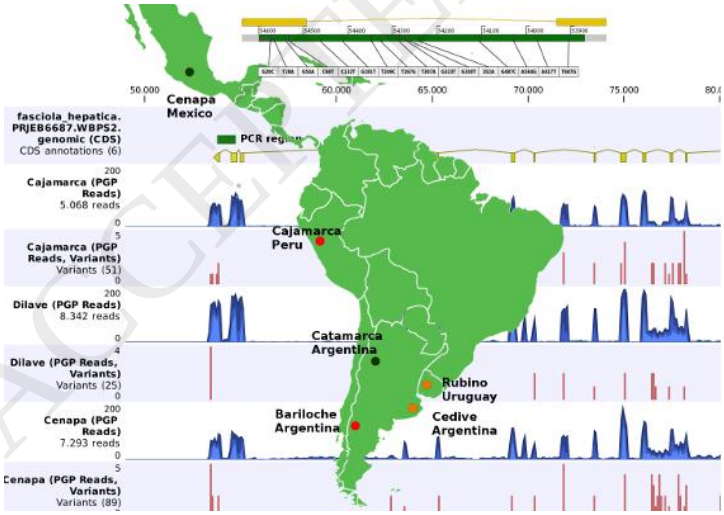
[c-] *Laboratorio de Parasitología, Centro de Investigación Veterinaria de Tandil (CIVETAN), CONICET, Facultad de Ciencias Veterinarias, UNCPBA, Tandil, Argentina.*

[1.] *Current address, Instituto de Investigaciones Biológicas Clemente Estable. IIBCE-MEC, Montevideo, Uruguay.*

E-mail addresses: MVS (mvsolana@vet.unicen.edu.ar), MFD (mafernandad@gmail.com), SS (vadyet.scarcella@gmail.com) SR (sradio91@gmail.com) PS (psmircich@fcien.edu.uy) SF (sfernand@vet.unicen.edu.ar) HS (hsolana@vet.unicen.edu.ar) JFT (jtort@fmed.edu.uy)

[#] Corresponding authors: Hugo Solana (hsolana@vet.unicen.edu.ar) and José F. Tort (jtort@fmed.edu.uy)

Graphical abstract



Highlights

- PGP variation was studied in 2 TCBZ resistant and 3 susceptible Latin American isolates
- No SNPs in the terminal domain of PGP could be associated with TCBZ resistance
- Several variable positions are detected along the 42 Kb of the FhPGP gene
- More detailed and extensive efforts are needed to identify markers of resistance

ABSTRACT

The use of Triclabendazole for controlling fasciolosis is compromised by increased drug resistance affecting livestock and humans. Although the mode of action of TCBZ is still unknown, putative candidates and markers of resistance have been advanced. A single nucleotide polymorphism (T687G) in *F. hepatica* PGP was proposed as marker of resistance in a small scale study of European susceptible and resistant flukes, but the association was not found in Australian samples. The T687G SNP was absent in more than 40 samples from 2 TCBZ-resistant and 3 susceptible isolates across Latin America here analyzed. While the American samples showed more variable SNPs than the previous ones, none of the SNPs detected showed a marked association with resistance. Analyzing the 42 kb of the FhPGP gene based on RNAseq data highlights that the variation has been underestimated, suggesting that more detailed efforts are needed in order to identify markers of resistance.

Keywords:

Fasciola hepatica, Latin America, Drug Resistance, Triclabendazole, P-glycoprotein, SNPs

Fasciolosis, is arguably one of the most widely distributed zoonotic diseases, with a major impact on livestock resulting in important economic losses estimated at thousands of millions of dollars globally [1]. The increased incidence of human cases has drawn attention to this neglected foodborne disease affecting approximately 50 million people worldwide [2]. In the Americas the disease is widespread in livestock, while is considered an important human foodborne infection in the Altiplano region of Bolivia and Peru [3]. Triclabendazole (TCBZ) is the drug of choice to combat this parasite based on its broad spectrum of activity against liver fluke of all ages down to 2 days post-infection in the definitive host [4]. However, the effectivity of this benzimidazolic drug has been compromised by reports of TCBZ resistance that first emerged in Australia, and were rapidly followed by cases in Europe, and in South America. More worrying, human cases reluctantly treated with TCBZ have been recently reported in Chile, and Peru (reviewed in [5].

Investigation of TCBZ resistance in the fluke in laboratories worldwide has resulted in the pursuit of a number of potential candidate genes and biological pathways [4,5]. The precise loci and, therefore, genes involved are still to be defined but a genome-wide approach is currently underway to identify the major genetic determinant of TCBZ resistance [6]. The increased cellular efflux of TCBZ in *F. hepatica* is related to ABC transporters as the P-glycoproteins (PGP) [4,7]. TCBZ-resistant isolates have been shown to process TCBZ more rapidly and their resistant phenotype can be reversed, under in vitro conditions, by the co-administration of inhibitors of P-glycoprotein (PGP) drug efflux pumps or drug detoxification pathways [4,8].

A previous study on European *Fasciola* isolates suggested that a single nucleotide polymorphism (SNPs) resulting in a PGP point mutation was associated with the resistant phenotype [9]. The same variation was not associated with resistance in Australian isolates [10]. In this study we analyzed the same terminal domain of *F. hepatica* PGP at the DNA and transcript levels in several samples with different status of resistance obtained from different locations across South America and compared them with those of the previous European and Australian studies. Our dataset consisted of individual flukes from 2 isolates defined as TCBZ-resistant from Bariloche, Argentina (7 samples) [11,12] and Cajamarca, Peru (10 DNA and 3 cDNA samples) [12,13], 7 flukes from an Argentinian isolate (Cedive, Chascomus) susceptible to TCBZ but resistant to Albendazole (ABZ) [12,14]. Beside this we included a TCBZ-susceptible isolate from Mexico (6 DNA and 7 cDNA samples), and 13 individuals flukes from field isolates from Catamarca, Argentina whose resistant status is unknown (Suppl. Figure 1). As an additional reference for this study, 5 adult flukes of the UK Cullompton susceptible isolate were also analyzed [12]. Considering this study and the two previous ones, the geographical width of the analysis of PGP variation is extended and the total amount of resistant and susceptible worms analyzed is roughly doubled (Suppl. Figure 1, Table 1).

We analyzed 16 variable positions in an amplimer of 830 bp, some of them shared with the previous studies and some novel. Eleven of the variable positions fall within the last intron, and the remaining 5 variable sites on the adjacent regions of the two final exons of PGP (Figure 1 B, C). All the American samples have an extra residue in a stretch of consecutive adenines within the intron that we noted here as position 352 indel. An extra base in this region was also reported in most of the samples from the Australian study [10], and is found in both available genome assemblies, and in 4 flukes of the European Cullompton isolate sequenced by us. Considering these differences we wonder

if the differences observed are due to a true indel or rather a differences in the base calling imposed by the presence of homopolymeric bases.

Conserved characteristics of all Latin American samples are the presence of a G residue instead of a T in position 267, while the site is variable in European and Australian samples. Also an enrichment in T in intronic position 617 is observed in the American dataset (25 homozygous and 7 heterozygotes of 44 samples) in relation to the other datasets (12/24 and 9/24 in Australian and European samples respectively) (Supplementary Table 1 A). In general we detected more variability in the Latin American samples, particularly with the more frequent presence of heterozygous positions (25/43 samples vs 7/20 [9] and 13/30 [10] in European and Australian samples respectively. Interestingly 11 individual flukes share the particular feature of being heterozygous at 3 concurrent intronic positions (C132T, G181T, G323T). The reason for this heterozygotic enrichment is not clear. Despite the wide geographical variation and resistant status of the American samples none of them presented the T687G SNP originally described as possibly associated with resistance, ruling out its association with the phenotype.

In order to gain further insights we organized the samples into haplotypes according to the various SNPs shared in the region amplified using DnaSP6 [15]. For this, we consider all the variants in the 2 previous studies [9,10] and those here obtained, resulting in 31 different haplotypes (Table 1). The American samples are the most diverse with 17 haplotypes present, compared to 10 in the European samples [9] and 7 in the Australian flukes [10]. The most frequent ones were shared, being haplotype FH02 [9] also present in the Australian study (referred as haplotype A) [10], but absent in America. Also highly abundant is haplotype 15 originally described as haplotype B by [10], that is shared between Australian and American samples. In any case no clear association emerged

between the diverse haplotypes and the resistant status, despite their differential distribution.

Since some of the intronic variations frequent in resistant flukes (particularly A617T) might result in a novel splicing acceptor site that would produce a longer final exon, we evaluated the presence of alternative splicing. RT-PCR performed on samples from susceptible and resistant isolates worms produced consistently a single band corresponding to the canonical predicted splicing site rather than two products differing by 45 bp, ruling out the possibility of alternative splicing (Supplementary Table 1 A).

Since we have recently performed a transcriptomic analysis of 3 American isolates with different drug susceptibility [16], we looked at variations detected in the FhPGP gene in this study. We retrieved Contig 2471 where FhPGP is found and all the reads mapped to it from the three isolates, namely Cajamarca, Peru (TCBZR), Cenapa, Mexico (TCBZS) and Rubino, Uruguay (TCBZS, ABZR) [12]. SNPs were called based on the reference genome [17] using the Basic Variant tools of the CLC software package using default parameters. Despite only 2 individual flukes were analyzed from each isolate, several variable sites were found distributed in the 42 kb that comprises the PGP gene, most of them within several of the 18 exons (Figure 1 C, Supplementary Table 2). The low amount of samples analyzed do not allow to infer selective presence of SNPs in resistant or susceptible isolates. However, it is worth mention that most variants are shared. These results highlight that sequence variability can be high within *F. hepatica* as it has been advanced by the genomic studies [17,18]. Consequently efforts to associate variant sites with particular phenotypes as drug resistance need to be very cautious before reporting.

In conclusion our data confirm that the T687G SNP in the PGP gene alleged as candidate for TCBZ resistance marker [9] is not detected in any of the American sequences analyzed compromising its proposed involvement in TCBZ resistance. Several other variable positions mapping within the last intron can be observed, but there is no evidence of altered splicing that might account for the resistant phenotype according to our RT-PCR experiments. Our study also highlight that the American isolates are slightly more variable at the PGP gene region under study than the European or Australian counterparts. This might be related to the largest dataset here analyzed and the fact that correspond to a widest geographical distribution of the Latin American samples. These results are also consistent with recent descriptions of genetic variability in *F. hepatica* in general [18]. Interestingly a recent study showed low genetic diversity in Peruvian flukes from Cajamarca wen analyzing mitochondrial ND1 gene and the ribosomal ITS [19]. While these results compromise the causal role of the previously indicated SNP, they do not exclude PGP as a candidate to resistance. Indeed, the region studied so far represent less than a tenth of the original protein, and our transcriptomic evidence [16] show that much variation exist along the 42 kb of the gene. Furthermore, the genomic sequences of *F.hepatica* have revealed multiple ABC transporters that might be playing a detoxifying role. The availability of genomic sequences allow now to perform more extensive studies analyzing in parallel variations in several genes across numerous individuals, opening new avenues for investigating variants associated with the resistance of *F. hepatica* to TCBZ.

Acknowledgments:

Funding: This work was supported by grant No 0865 awarded to HS, from the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT), Argentina. MVS is a recipient of a PhD scholarship from CONICET, Argentina. We would like to thank Drs Pedro Ortiz (Univ. Nacional de Cajamarca, Peru), Fermin Olaechea (INTA-Bariloche,

Argentina), Estephan Miranda (CENAPA, Mexico) and Valeria Gayo (DILAVE Rubino, Uruguay) for providing samples for this study.

References

[1] K. Mehmood, H. Zhang, A.J. Sabir, R.Z. Abbas, M. Ijaz, A.Z. Durrani, M.H. Saleem, M. Ur Rehman, M.K. Iqbal, Y. Wang, H.I. Ahmad, T. Abbas, R. Hussain, M.T. Ghori, S. Ali, A.U. Khan, J. Li, A review on epidemiology, global prevalence and economical losses of fasciolosis in ruminants, Microb. Pathog. 109 (2017) 253–262. doi:10.1016/j.micpath.2017.06.006.

[2] M. Nyindo, A.-H. Lukambagire, Fascioliasis: An ongoing zoonotic trematode infection, Biomed Res. Int. 2015 (2015) 1–8. doi:10.1155/2015/786195.

[3] C. Carmona, J.F. Tort, Fasciolosis in South America: epidemiology and control challenges, J. Helminthol. 91 (2017) 99–109. doi:10.1017/S0022149X16000560.

[4] G.P. Brennan, I. Fairweather, A. Trudgett, E. Hoey, McCoy, M. McConville, M. Meaney, M. Robinson, N. McFerran, L. Ryan, C. Lanusse, L. Mottier, L. Alvarez, H. Solana, G. Virkel, P.M. Brophy, Understanding triclabendazole resistance., Exp. Mol. Pathol. 82 (2007) 104–109. doi:10.1016/j.yexmp.2007.01.009.

[5] J.M. Kelley, T.P. Elliott, T. Beddoe, G. Anderson, P. Skuce, T.W. Spithill, Current threat of triclabendazole resistance in *Fasciola hepatica*, Trends Parasitol. 32 (2016) 458–469. doi:10.1016/j.pt.2016.03.002.

[6] J. Hodgkinson, K. Cwiklinski, N.J. Beesley, S. Paterson, D.J.L. Williams, Identification of putative markers of triclabendazole resistance by a genome-wide analysis of genetically recombinant *Fasciola hepatica*, Parasitology. 140 (2013) 1523–1533. doi:10.1017/S0031182013000528.

[7] L.I. Alvarez, H.D. Solana, M.L. Mottier, G.L. Virkel, I. Fairweather, C.E. Lanusse,

Altered drug influx/efflux and enhanced metabolic activity in triclabendazole-resistant liver flukes., Parasitology. 131 (2005) 501–510. doi:10.1017/S0031182005007997.

[8] M. Meaney, J. Savage, G.P. Brennan, E. Hoey, A. Trudgett, I. Fairweather, Increased susceptibility of a triclabendazole (TCBZ)-resistant isolate of *Fasciola hepatica* to TCBZ following co-incubation in vitro with the P-glycoprotein inhibitor, R(+)-verapamil., Parasitology. 140 (2013) 1287–1303. doi:10.1017/S0031182013000759.

[9] R. Wilkinson, C.J. Law, E.M. Hoey, I. Fairweather, G.P. Brennan, A. Trudgett, An amino acid substitution in *Fasciola hepatica* P-glycoprotein from triclabendazole-resistant and triclabendazole-susceptible populations, Mol. Biochem. Parasitol. 186 (2012) 69–72. doi:10.1016/j.molbiopara.2012.08.008.

[10] T.P. Elliott, T.W. Spithill, The T687G SNP in a P-glycoprotein gene of *Fasciola hepatica* is not associated with resistance to triclabendazole in two resistant Australian populations, Mol. Biochem. Parasitol. 198 (2014) 45–47. doi:10.1016/j.molbiopara.2014.11.006.

[11] F. Olaechea, V. Lovera, M. Larroza, F. Raffo, R. Cabrera, Resistance of *Fasciola hepatica* against triclabendazole in cattle in Patagonia (Argentina)., Vet. Parasitol. 178 (2011) 364–366. doi:10.1016/j.vetpar.2010.12.047.

[12] J. Canevari, L. Ceballos, R. Sanabria, J. Romero, F. Olaechea, P. Ortiz, M. Cabrera, V. Gayo, I. Fairweather, C. Lanusse, L. Alvarez, Testing albendazole resistance in *Fasciola hepatica*: validation of an egg hatch test with isolates from South America and the United Kingdom, J. Helminthol. 88 (2014) 286–292. doi:10.1017/S0022149X13000163.

[13] P. Ortiz, S. Scarcella, C. Cerna, C. Rosales, M. Cabrera, M. Guzmán, P. Lamenza, H. Solana, Resistance of *Fasciola hepatica* against Triclabendazole in cattle in Cajamarca (Peru): A clinical trial and an in vivo efficacy test in sheep, Vet. Parasitol. 195 (2013) 118–121. doi:10.1016/j.vetpar.2013.01.001.

[14] R. Sanabria, L. Ceballos, L. Moreno, J. Romero, C. Lanusse, L. Alvarez, Identification of a field isolate of *Fasciola hepatica* resistant to albendazole and susceptible to triclabendazole., Vet. Parasitol. 193 (2013) 105–10. doi:10.1016/j.vetpar.2012.11.033.

[15] J. Rozas, A. Ferrer-Mata, J.C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado, S.E. Ramos-Onsins, A. Sánchez-Gracia, DnaSP 6: DNA sequence polymorphism analysis of large data sets, Mol. Biol. Evol. (2017) 1–4. doi:10.1093/molbev/msx248.

[16] S. Radio, S. Fontenla, V. Solana, A.C. Matos Salim, F.M.G. Araújo, P. Ortiz, C. Hoban, E. Miranda, V. Gayo, F.S.-M. Pais, H. Solana, G. Oliveira, P. Smircich, J.F. Tort, Pleiotropic alterations in gene expression in Latin American *Fasciola hepatica* isolates with different susceptibility to drugs, Parasit. Vectors. 11 (2018) 56. doi:10.1186/s13071-017-2553-2.

[17] K. Cwiklinski, J.P. Dalton, P.J. Dufresne, J. La Course, D.J. Williams, J. Hodgkinson, S. Paterson, The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution, Genome Biol. 16 (2015) 71. doi:10.1186/s13059-015-0632-2.

[18] N.J. Beesley, D.J.L. Williams, S. Paterson, J. Hodgkinson, *Fasciola hepatica* demonstrates high levels of genetic diversity, a lack of population structure and high gene flow: possible implications for drug resistance, Int. J. Parasitol. 47 (2017) 11–20. doi:10.1016/j.ijpara.2016.09.007.

[19] M. Ichikawa-Seki, P. Ortiz, M. Cabrera, C. Hobán, T. Itagaki, Molecular characterization and phylogenetic analysis of *Fasciola hepatica* from Peru, Parasitol. Int. 65 (2016) 171–174. doi:10.1016/j.parint.2015.11.010.

[20] S.N. McNulty, J.F. Tort, G. Rinaldi, K. Fischer, B.A. Rosa, P. Smircich, S. Fontenla, Y. Choi, R. Tyagi, K. Hallsworth-Pepin, V.H. Mann, L. Kammili, P.S. Latham, N. Dell'Oca, F. Dominguez, C. Carmona, P.U. Fischer, P.J. Brindley, M. Mitreva, Genomes

of *Fasciola hepatica* from the Americas reveal colonization with *Neorickettsia* endobacteria related to the agents of Potomac Horse and Human Sennetsu Fevers, PLOS Genet. 13 (2017) e1006537. doi:10.1371/journal.pgen.1006537.

Figure Legends

Figure 1. (A) SNPs detected by genomic and RT PCR in different South American isolates. The PCR primer set (forward: TTGGTGTTGTATCGCAGGAA; reverse: AGCCGAAGTAGCTTCATCCA) amplified the second nucleotide binding domain of the PGP gene as described [9]. An amplification product of 830 bp (green bar in figure) was obtained when amplifying from genomic DNA, due to the presence of a 569 bp intron (yellow line, with boxes representing adjacent exons), and a shorter product of the expected size was obtained from RT-PCR reactions. Fragments were sequenced at Macrogen, Korea. Sequences were CAP aligned to the reference PGP gene in the two *F. hepatica* genome assemblies available [17,20]. In order to facilitate comparisons, we numbered variable positions according to PCR fragment position following the original report [9], and their relative position is indicated in the figure. Haplotypes were derived using DNASP6 [15]. Since we don't know the phases in heterozygous positions, those were encoded with IUPAC consensus bases (i.e. R for G/A) and used as input for the program to estimate the haplotypes. Similarly the data from previous studies [10] were included for comparison. The genotypes of all the samples are presented in Supplementary Table 1 and the resulting haplotypes are reported in Table 1. Sequences were deposited at GenBank (accessions MH267603-MH267660) **(B). SNPs detected in PGP gene by RNAseq in three different South American isolates.** RNA extracted from two individual worms each of the Cajamarca, Peru (TCBZR-ABZR), Rubino, Uruguay (TCBZS ABZR) and Cenapa, Mexico (TCBZS ABZS) isolates were used to generate

Illumina PE- RNAseq libraries. The resulting sequencing data was quality trimmed and mapped to the *F.hepatica* genome [17] with the CLC Genomics Workbench package (Qiagen, Aarhus A/S) as described [16]. For variant calling, we retrieved the contig 2471 where FhPGP is found and all the reads that mapped to it from the three isolates analyzed. SNPs were called using the Variant tools of the CLC software package with default parameters (minimum coverage 10, minimum frequency 20%). The scheme represents the genomic structure of the FhPGP gene (BN1106_s2471B000078, on reverse strand) depicted in yellow in the top panel, with the region analyzed by PCR indicated (green box). RNAseq coverage on the Cajamarca (TCBZR, red dot) Cenapa, (TCBZS, green dot) and Rubino (ABZR, TCBZS, orange dot) isolates are indicated by blue peaks in independent tracks. SNPs variants detected along the whole coding region in each isolate are indicated as red bars at the bottom of each read map and detailed in Supplementary Table 2. Total reads and variants counted on the region are indicated on the track labels. RNAseq data was deposited as raw data at the SRA database under accession PRJNA339158.

Table 1. Haplotypes found in this study (shaded) and in previous reports.

Haplotypes generated by DnaSP6 [15] based on data obtained in this study (gray shaded) and in previous studies [9, 10]. Haplotype 1 of Wilkinson et al. [9] was taken as reference (Haplotype H02 of this study), and the variants detected are resumed in 2nd column, and detailed by position in the following columns. Those SNPs corresponding to exons are shaded. The total count of each haplotype, and their presence in sensitive and resistant individuals from different populations is resumed in the second part of the table (#). Some alternative haplotypes for genotypes already described were predicted in this study and are indicated (*). The correlation of the haplotypes names between this study and the previous ones is indicated in the last column.

HAPLOTYPE	Single Nucleotide Polymorphisms (SNPs)																		Haplotype count #										Haplotype in previous studies
	Variants count	Variant Positions detected																	Total	Europe			Australia		America			Genome		
		G20C	T28A	G54A	C60T	C132T	G181T	T249C	T267G	T307A	G323T	G330T	352A	G487C	A564G	A617T	T687G			Sens	RES		Sens	RES	Sens	RES	unk	unk		
H_02	0	G	T	G	C	C	G	T	T	T	G	G	-	G	A	A	T		44	15	8		5	16						Wilkinson 1, Elliot A
H_01	2								G				A						4				1				1	2		Elliot E
H_03	4		A	A	T								-						2	2										Wilkinson 4
H_04	5						T		G		T		-		G				1	1										Wilkinson 2*
H_05	7		A	A			T		G		T		-				G		1	1										Wilkinson 3*
H_06	2		A										-						1	1										Wilkinson 5
H_07	3		A							A			-						1		1									Wilkinson 9*
H_08	6	C	A				T			A			-				G		1		1									Wilkinson 10*
H_09	3								G				-			T			6		6									Wilkinson 8
H_10	4						T		G				-			T			1		1									Wilkinson 6
H_11	6						T		G		T		-			T	G		2		2									Wilkinson 7
H_12	4						T		G		T		-						1		1									
H_13	4							C	G				A			T			3	3										
H_14	4		A						G				A			T			7	3						4				
H_15	3								G				A			T			48	2			6	6	7	10	17			Elliot B
H_16	4								G				A	C		T			4				4							Elliot D
H_17	4						T		G				A			T			8				2	2	1	2	1			Elliot C
H_19	4					T	T		G				A						2									2		
H_20	4						T		G		T		A						1							1				
H_21	5					T	T		G		T		A						4							1	3			
H_22	5								G		T		A		G	T			3							2	1			
H_23	6					T	T		G		T		A		G				9						3	3	3			

H_24	3								G				A		G				15						8	7					
H_25	4					T			G				A			T			2						1	1					
H_26	5					T	T		G				A			T			3						1	2					
H_27	4								G			T	A			T			2						1	1					
H_28	6					T	T		G		T		A			T			1						1						
H_29	5						T		G		T		A			T			1						1						
H_30	5								G	C	T		A			T			1						1						
H_31	4								G			T	A		G				1						1						