Forensic Population Genetics – Letter to the Editor

Analysis of 17 STR data on 5362 southern Portuguese individuals – an update on reference database

Raquel Cabezas Silva [a,b]

[a] National Institute of Legal Medicine and Forensic Sciences, Portugal

[b] University of Lisbon, Faculty of Medicine, Portugal

Teresa Ribeiro [a,b]

[a] National Institute of Legal Medicine and Forensic Sciences, Portugal

[b] CENCIFOR - Forensic Sciences Centre, Portugal

Isabel Lucas [a]

[a] National Institute of Legal Medicine and Forensic Sciences, Portugal

Maria João Porto [a,b]

[a] National Institute of Legal Medicine and Forensic Sciences, Portugal

[b] CENCIFOR—Forensic Sciences Centre, Portugal

Jorge Costa Santos [a,b,c]

[a] National Institute of Legal Medicine and Forensic Sciences, Portugal

[b] CENCIFOR - Forensic Sciences Centre, Portugal

[c] University of Lisbon, Faculty of Medicine, Portugal

Paulo Dario [a,b,c,d,*]

[a] National Institute of Legal Medicine and Forensic Sciences, Portugal

[b] University of Lisbon, Faculty of Sciences, Portugal

[c] CENCIFOR - Forensic Sciences Centre, Portugal

[d] CESAM – Centre for Environmental and Marine Studies, Lisboa, Portugal

Authors contacts:

Raquel Silva, e-mail: raquel.csiilva@gmail.com

Teresa Ribeiro, e-mail: tribeiro@dlinml.mj.pt

Isabel Lucas, e-mail: ilucas@dlinml.mj.pt

Maria João Porto, e-mail: mariajoao.porto@dcinml.mj.pt

Jorge Costa Santos, e-mail: jsantos@dlinml.mj.pt

Paulo Dario, e-mail: pdario@dlinml.mj.pt

*Corresponding author at: Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. – Delegação do Sul. Rua Manuel Bento de Sousa n.º 3. 1169-201 Lisboa, Portugal. Tel.: +351 218 811 866; Fax: +351 218 809 958; E-mail address: pdario@dlinml.mj.pt (P. Dario).

Forensic Population Genetics – Letter to the Editor

Analysis of 17 STR data on 5362 southern Portuguese individuals – an update on reference database

ABSTRACT

The main objective of this work consisted of the updating of allele frequencies and other relevant forensic parameters for the 17 autosomal STR loci provided by the combination of the two types of kits used routinely in our laboratory casework: AmpF*l*STR Identifiler® and the Powerplex® 16 Systems. This aim was of significant importance, given that the last study on these kits within the southern Portuguese population dates back to 2006, and, as a consequence, it was necessary to correct the deviation caused by population evolution over the last ten years so that they might be better applied to our forensic casework. For this reason genetic data from 5362 unrelated Caucasian Portuguese individuals from the south of Portugal who were involved in paternity testing casework from 2005 to 2014 was used. Of all the markers, TPOX proved to be the least polymorphic, and Penta E the most. Secondly, this up-to-date southern Portuguese population was compared not only with the northern and central Portuguese populations, but also with that of southern Portugal in 2006, along with populations from Spain, Italy, Greece, Romania, Morocco, Angola and Korea in order to infer information about the relatedness of these respective populations, and the variation of the southern Portuguese population over time.

Keywords: Identifiler®; Powerplex® 16; Population data update; Population genetics; Portugal.

Dear Editor,

As we all know, using proper allele frequencies and statistical data representative of the population in study is of great importance in forensic casework. In order to obtain this information, populations of interest are studied and their data published so that the scientific community can access and use that information when necessary; this journal being a good source of such kinds of work. However, over time, population parameters can undergo certain deviations from their original state. Studies that were representative of the time in which they were conducted, rather like a photograph taken at that particular time, no longer accurately reflect the present conditions. Genetic variation occurs due to *phenomena* such as migration or genetic drift, both of which can affect population statistical parameters, making it imperative to keep databases updated as much as possible.

The last study on AmpF*l*STR Identifiler® (Applied Biosystems) and Powerplex® 16 System (Promega Corporation) in the southern Portuguese population was published in 2006 [1]. Since these, and similar kits are still the main ones used by this laboratory in both paternity testing and forensic casework, it was necessary to update our reference database.

In order to do that, we performed a retrospective study using the genotypic data of 5362 unrelated, Caucasian, southern Portuguese individuals involved in paternity testing casework from 2005 to 2014. All of these individuals gave their

49 informed consent at the time that they were sampled for blood and saliva. Blood
50 samples were analyzed with AmpF/STR Identifiler® or AmpF/STR Identifiler®
51 Plus, and buccal swabs were analyzed with Powerplex® 16 or Powerplex® 16
52 HS System over the ten year period. During laboratory analysis, internal
53 procedures consisting of half volume modified manufacturer's instructions,
54 followed by capillary electrophoresis in Applied Biosystems 3130xl or 3130
55 Genetic Analyzers were used. These kits, combined, included 17 autosomal
56 STR loci: CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179,
57 D13S317, D16S539, D18S51, D19S433, D21S11, FGA, TH01, TPOX, vWA,
58 Penta D and Penta E. All genotypic information was then collected from the .fst
59 files stored in our genetic analyzer backup systems, and visually confirmed with
60 the electrophoregrams obtained at the time they were first generated for
61 paternity testing purposes. The information of both kits was combined to obtain
62 the complete 17 loci genetic information that was then compared to the
63 information present in the reports. Following this, the data was filtered in order
64 to extract only the pertinent individuals - Caucasian unrelated Portuguese - and
65 was anonymized for the present study.

66 Allele frequencies, observed heterozygosity (Ho), expected heterozygosity (He),
67 Hardy–Weinberg equilibrium (HWE) and Linkage Disequilibrium (LD) were
68 estimated using Arlequin v3.5.1.2 software [2]. The calculations of population
69 pairwise genetic distances (F_{ST}) between the contemporary southern
70 Portuguese population and the others were also performed with Arlequin v3.5.
71 [2]. The populations compared in this study were the northern Portuguese [3]
72 and central Portuguese [4] and also the 2006 southern Portuguese, [1]
73 alongside those from Spain [5], Italy [6], Greece [7], Romania [8], Morocco [9],

Angola [10] and Korea [11]. With Arlequin's results a phylogram was constructed using Molecular Evolutionary Genetics Analysis v.6.06 software [12] applying a neighbour-joining methodology. Statistical parameters of forensic interest such as power of discrimination (PD), power of exclusion (PE), polymorphic information content (PIC), typical paternity index (TPI), and matching probability (MP) were calculated with PowerStats v1.2 [13] spreadsheet, modified by the authors in order to support and manage the large amount of samples. Minimum allele frequencies (MAF) were calculated as 5/2N. Allele frequencies along with statistical and forensic parameters are presented in Supplementary Table I. The average level of genetic diversity (He) was 0.801 and the most variable loci were: D18S51, Penta E, FGA and D2S1338 with 16 or more alleles each, and a He of over 85%. TPOX proved to be the least polymorphic marker and Penta E the most. Power of discrimination (PD) ranged from 0.823 (TPOX) to 0.977 (Penta E), with the combined PD equal to 0.999999999999999999997. Power of exclusion ranged from 0.344 (TPOX) to 0.756 (Penta E), with the combined PE equal to 0.99999989. MAF values varied between 0.000466 and 0.000473, depending on loci. There were four loci with statistically significant deviations from HWE ($p > 0.05$). After applying the Bonferroni's correction ($p > 0.0029$) there are still three loci with statistically significant deviations from HWE (D18S51, Penta D and TPOX). The most problematic marker was TPOX which had a *p* value of zero. This is the least polymorphic locus with the highest frequency of homozygosity in our study, resulting in a significant HWE deviation, probably caused by their heterozygote deficiency. From all the CODIS markers, TPOX shows the least variation between individuals [14]. The same happens here to this marker. Although, *p*

values of zero indicate deviations from HWE, these occur because we are not dealing with perfect populations subject to Hardy-Weinberg principles. Because of this, some studies can be found where this kind of observation occurs [15,16], some in a large number of loci [17]. In fact, in the three affected markers the Ho was lower than the He, which means theoretically, we would expect a higher heterozygosity, which was not the case. In nature, phenomena like inbreeding, for example, may cause deviations from HWE due to the decrease of random mating, increasing homozygosity. LD was evaluated using shuffling test for all possible combinations between loci. Twenty pairs of loci presented significant LD ($p<0.05$) among 136 pairwise comparisons. After applying Bonferroni correction only two pairs of loci exhibited significant LD: the pair Penta E and Penta D and the pair D2S1338 and D19S433, both with $p=0.00000$. Because these are not in the same chromosomes, these loci have been considered genetically unlinked. Therefore, all 17 markers could be treated as independent loci at the population level, which means no LD was detected in the studied loci in the southern Portuguese population.

Pairwise genetic distances values between populations from Portugal (southern, northern and central, as well as those from the 2006 southern Portuguese study), together with Spain, Italy, Greece, Romania, Morocco, Angola, and Korea are presented in Supplementary Table II and displayed in Fig. 1 in phylogram form. As expected, the population from Korea and Angola diverged the most from that of southern Portugal, with F_{ST} values of 0.01975 and 0.0114, respectively. It is important to note that the variation of the southern Portuguese Population was minimal with a negative F_{ST} value, which can be interpreted as the absence of population differentiation.

This laboratory is ISO/IEC 17025:2005 accredited and participates in the collaborative quality control and proficiency testing exercises of the Spanish and Portuguese Working Group (GHEP) from ISFG. This letter follows the guidelines for publication of population data as requested by the journal [18,19]. INMLCF – The National Institute of Legal Medicine and Forensic Sciences, Portugal, financed this work.

References

[1] C. Vieira-Silva, C. Cruz, T. Ribeiro, R. Espinheira, South Portugal population genetic analysis with 17 loci STRs, Int. Congr. Ser. 1288 (2006) 367–368.

[2] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567.

[3] M.F. Pinheiro, L. Cainé, L. Pontes, D. Abrantes, G. Lima, M.J. Pereira, et al., Allele frequencies of sixteen STRs in the population of Northern Portugal., Forensic Sci. Int. 148 (2005) 221–3.

[4] V. Lopes, A. Serra, J. Gamero, L. Sampaio, F. Balsa, C. Oliveira, et al., Allelic frequency distribution of 17 STRs from Identifiler and PowerPlex-16 in Central Portugal area and the Azores archipelago., Forensic Sci. Int. Genet. 4 (2009) e1–7.

[5] O. García, J. Alonso, J.A. Cano, R. García, G.M. Luque, P. Martín, et al., Population genetic data and concordance study for the kits Identifiler, NGM,

PowerPlex ESX 17 System and Investigator ESSplex in Spain., Forensic Sci. Int. Genet. 6 (2012) e78–9.

[6] F. Brisighelli, C. Capelli, I. Boschi, P. Garagnani, M. V Lareu, V.L. Pascali, et al., Allele frequencies of fifteen STRs in a representative sample of the Italian population., Forensic Sci. Int. Genet. 3 (2009) e29–30.

[7] P. Sánchez-Diz, P.G. Menounos, A. Carracedo, I. Skitsa, 16 STR data of a Greek population, Forensic Sci. Int. Genet. 2 (2008) e71–2.

[8] A. Anghel, A. Enache, E. Seclaman, G. Gruin, S. Ursoniu, A. Alexa, et al., Genetic polymorphism data on 15 autosomal STR markers in a Western Romanian population sample., Leg. Med. (Tokyo). 16 (2014) 238–40.

[9] K. Bentayebi, F. Abada, H. Ihzmad, S. Amzazi, Genetic ancestry of a Moroccan population as inferred from autosomal STRs., Meta Gene. 2 (2014) 427–38.

[10] M.M. Melo, M. Carvalho, V. Lopes, M.J. Anjos, A. Serra, D.N. Vieira, et al., Genetic study of 15 STRs loci of Identifiler system in Angola population, Forensic Sci. Int. Genet. 4 (2010) e153–e157.

[11] S.B. Hong, S.H. Kim, K.C. Kim, M.H. Park, J.Y. Lee, J.M. Song, et al., Korean population genetic data and concordance for the PowerPlex ® ESX 17, AmpFlSTR Identifiler®, and PowerPlex® 16 systems, Forensic Sci. Int. Genet. 7 (2013) e47–51.

[12] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.

[13] A. Tereba, Tools for analysis of population statistics, Profiles in DNA 3, Promega Corporation, USA, 1999, pp. 14–16.

[14] J.M. Butler, Genetics and genomics of core short tandem repeat loci used in human identity testing, J. Forensic Sci. 51 (2006) 253–265.

[15] C.R. Hill, D.L. Duewer, M.C. Kline, M.D. Coble, J.M. Butler, U.S. population data for 29 autosomal STR loci., Forensic Sci. Int. Genet. 7 (2013) e82–3.

[16] M. Soták, E. Petrejčíková, A. Bôžiková, J. Bernasovská, I. Bernasovský, A. Sovičová, et al., Population database of 17 autosomal STR loci from the four predominant Eastern Slovakia regions., Forensic Sci. Int. Genet. 5 (2011) 262–3.

[17] M. V. Albeza, A. Picornell, N. Acreche, C. Tomas, J.A. Castro, M.M. Ramon, Genetic variability at 14 STR loci in the Puna population of north western Argentina, Int. J. Legal Med. 116 (2002) 126–132.

[18] V.A. Stepanov, A. V. Melnikov, A.Y. Lash-Zavada, V.N. Kharkov, S.A. Borinskaya, T. V. Tyazhelova, et al., Genetic variability of 15 autosomal STR loci in Russian populations, Leg. Med. 12 (2010) 256–258.

[19] A. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, et al., New guidelines for the publication of genetic population data., Forensic Sci. Int. Genet. 7 (2013) 217–20.

[20] J.M. Carracedo, A. Gusmão, W. Linacre, L. Roewer, et al., Update of the guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. 10 (2014) A1–A2.

193

194 20 October 2015

195

196

197 **Figure Legends**

198

199

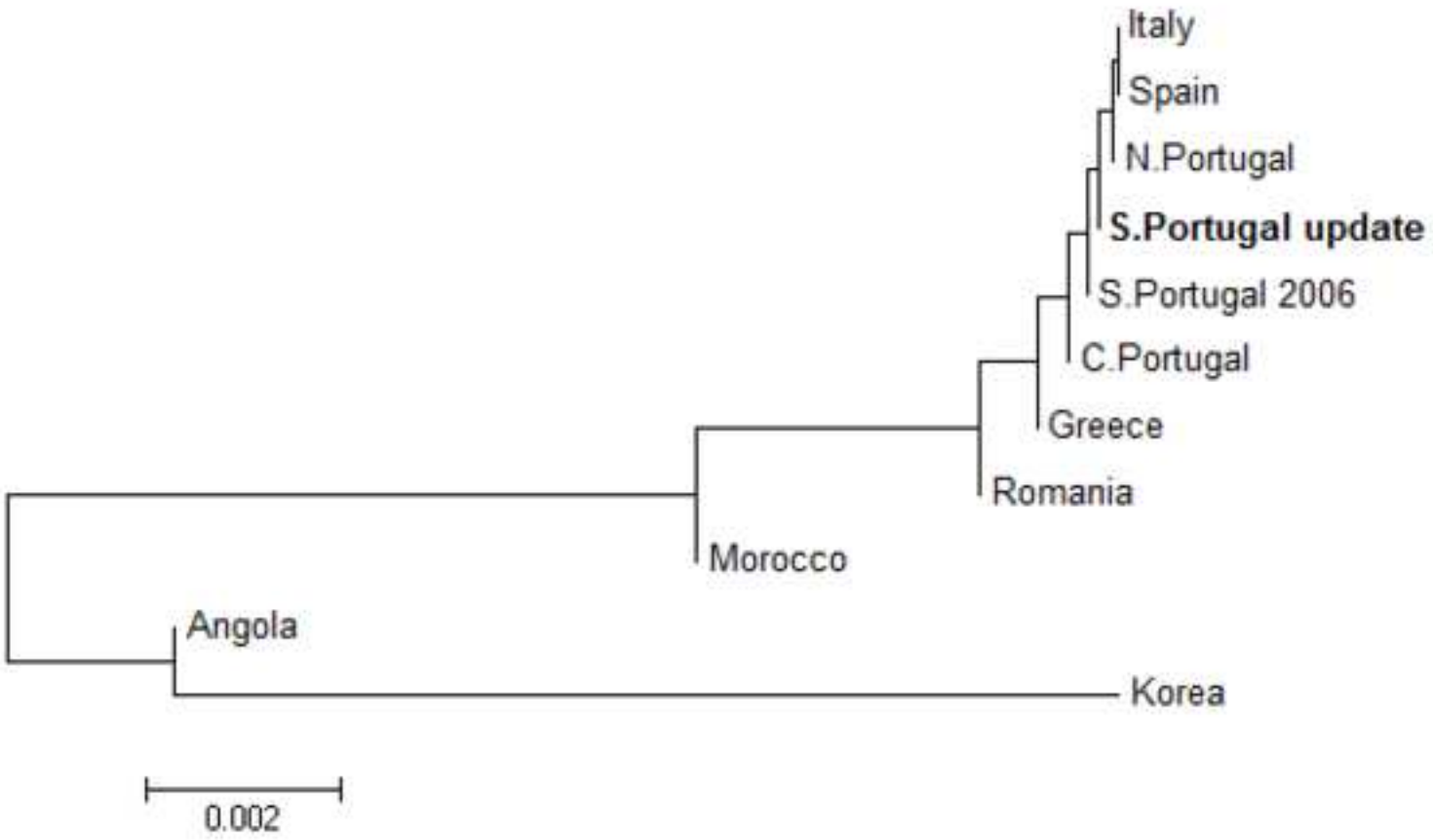
200

201 Fig. 1. Neighbor-joining phylogram for the 17 combined Identifiler® and
202 Powerplex® 16 autosomal STR loci from South (updated and 2006 study), North
203 and Central Portugal, Spain, Italy, Greece, Romania, Morocco, Angola, and
204 Korea populations.

205

Fig. 1. Neighbor-joining phylogram for the 17 combined Identifil

Click here to download high resolution image



Supplementary Table I - Allele frequencies and statistical and

Click here to download e-component: Supplementary Table I - Analysis of 17 STR.xlsx

Supplementary Table I - Allele frequencies and statistical and forensic parameters for the 17

	D3S1358	TH01	D21S11	D18S51	PentaE	D5S818	D13S317	D7S820
2.2								
3		0.0001						
3.2								
4		0.0007						
4.2					0.0002			
5		0.0009			0.0620		0.0001	
6		0.2079			0.0012		0.0002	
6.3								0.0005
7		0.1716			0.1418	0.0004	0.0001	0.0169
7.3								
8		0.1334			0.0179	0.0086	0.1361	0.1552
8.3		0.0003						
9		0.1968		0.0002	0.0132	0.0308	0.0563	0.1286
9.1								0.0003
9.3		0.2753						
10	0.0001	0.0131		0.0137	0.0852	0.0611	0.0495	0.2717
10.1								0.0001
10.2				0.0005				
10.3		0.0001						
11	0.0011			0.0091	0.1311	0.3405	0.3275	0.2274
11.2								
12	0.0027			0.1402	0.2032	0.3767	0.2772	0.1623
12.1						0.0002		
12.2								
13	0.0029			0.1302	0.1127	0.1710	0.1115	0.0311
13.2				0.0001				
14	0.1049			0.1429	0.0571	0.0104	0.0405	0.0054
14.2				0.0002				
14.4					0.0001			
15	0.2645			0.1467	0.0454	0.0005	0.0009	0.0005
15.2				0.0001				
15.4					0.0001			
16	0.2509			0.1585	0.0410			
16.1								
16.2								
17	0.2082			0.1121	0.0363			
17.2								
18	0.1528			0.0651	0.0184			
18.1				0.0001				
18.2								
18.3	0.0001							
19	0.0113			0.0421	0.0174			
19.2								
20	0.0005			0.0242	0.0081			
21				0.0088	0.0054			
21.2								
22				0.0039	0.0013			

Supplementary Table II - Population pairwise

Click here to download e-component: Supplementary Table II - Population pairwise .xlsx

Supplementary Table II - Population pairwise FSTs between the populations from Portug

	S.Portugal upd	North Portugal	Central Portug	S.Portugal 200	Spain
S.Portugal updated	x				
North Portugal	-0.00146	x			
Central Portugal	-0.00035	-0.00095	x		
S.Portugal 2006	-0.00022	-0.00095	-0.00035	x	
Spain	-0.00017	-0.00138	-0.00026	-0.00029	x
Italy	0.00008	-0.00096	0.00012	0.00016	-0.00197
Greece	-0.00091	-0.00185	-0.00074	-0.00051	0.00009
Romania	-0.00120	-0.00104	-0.00142	-0.00117	-0.00167
Morocoo	0.00020	-0.00012	0.00026	0.00057	0.00262
Angola	0.01104	0.01291	0.01054	0.01042	0.01857
Korea	0.01975	0.02777	0.01842	0.01906	0.02848

Supplementary data - Analysis of 17 STR data on 5362 South Portu

Click here to download Additional Files: Supplementary data - Analysis of 17 STR data .xlsx

ID	D3S1358		TH01		D21S11		D18S51		PentaE		D5S818		D13S317		D7S820	
0001	15	16	6	8	29	30	14	16			12	12	12	12	8	11
0002	14	18	9	9.3	29	30	14	16	10	12	9	12	11	11	10	11
0003	14	17	7	9.3	28	31.2	21	22	7	13	11	13	9	12	8	10
0004	15	16	6	8	29	32.2	16	23	13	13	11	12	12	12	10	11
0005	16	17	8	9	30	31.2	17	18	5	7	11	13	10	12	7	8
0006	14	16	8	9.3	29	31.2	15	17	12	12	11	13	11	12	10	12
0007	15	15	9.3	9.3	29	29	12	12	11	14	11	12	11	11	10	12
0008	14	15	6	9	31.2	31.2	17	19	5	12	10	13	11	11	11	11
0009	18	18	6	7	29	30	11	12	11	11	12	13	11	12	10	11
0010	14	18	9	9.3	30	30.2	13	14	12	12	13	13	11	13	10	13
0011	15	18	7	7	30	33.2	12	14	5	18	11	12	11	12	8	11
0012	16	17	6	6	27	31	16	17	12	20	12	12	11	12	10	11
0013	16	16	8	9	29	30.2	12	17	10	11	10	11	8	12	10	12
0014	18	19	6	8	27	30	16	16	12	20	11	12	8	11	10	10
0015	15	16	6	6	25	31.2	14	16	11	12	10	12	12	14	10	10
0016	15	15	6	9	30	31	12	14	11	15	11	12	11	11	8	9
0017	15	15	8	9.3	30	32	15	17	13	17	12	13	11	12	8	11
0018	16	18	9.3	9.3	28	32	14	18	8	11	11	12	12	14	11	12
0019	15	16	7	9.3	29	31.2	16	20			12	13	8	8	8	12
0020	17	17	6	9	30	31	13	16	7	16	10	12	8	11	11	11
0021	14	19	7	9.3	28	31.2	13	16	11	12	11	12	8	11	10	11
0022	16	18	6	8	28	30	15	16	12	12	11	12	8	11	8	10
0023	16	18	7	9.3	31.2	32.2	17	17	5	12	12	12	11	11	9	12
0024	14	15	6	9.3	28	29	12	17	12	13	10	13	11	12	8	10
0025	17	18	6	8	29	30	16	18	9	13	12	12	11	12	9	10
0026	14	18	8	9.3	30	30	13	15			10	12	8	8	10	11
0027	15	15	6	7	28	29	12	15	10	10	13	13	8	11	10	11
0028	16	17	9	9.3	29	30	12	15	5	13	12	12	11	12	9	11
0029	14	16	6	8	29	29	13	15	5	12	11	12	11	14	11	13
0030	16	18	9	9.3	29	31.2	12	18	5	7	11	12	11	12	11	11
0031	17	18	6	8	29	31.2	12	18	11	12	11	12	12	14	10	12
0032	13	14	9	9	29	31.2	14	16	11	19	11	11	8	12	9	11
0033	16	16	7	9.3	29	31	10	15	12	12	12	13	11	12	8	10
0034	17	18	7	9	30	30	11	16	11	20	12	13	10	11	10	12
0035	15	15	8	9	29	29	14	18	12	13	12	12	8	11	10	11
0036	16	18	8	9.3	28	33.2	15	16	13	17	12	12	9	14	9	11
0037	17	19	6	7	29	32.2	14	16	10	15	11	12	11	11	10	12
0038	17	17	7	9	29	33.2	13	19	11	12	11	12	11	11	8	9
0039	16	18	8	9.3	30	30	13	17	10	13	9	11	9	11	10	13
0040	15	18	9.3	9.3	30	31	13	15	11	15	11	12	11	12	10	12
0041	15	18	9	9	27	32.2	14	15			11	13	11	11	10	11
0042	15	18	6	9.3	30.2	32.2	12	13	11	13	12	12	9	12	8	12
0043	15	16	6	9.3	31	32.2	16	18	10	10	10	12	8	11	9	10
0044	14	18	9.3	9.3	28	31.2	14	17			12	12	12	12	8	12
0045	16	17	7	7	29	33.2	14	19	12	14	11	12	9	11	8	9
0046	15	17	6	7	28	29	14	15	10	13	11	11	11	12	10	10