



HAL
open science

Hepatitis C virus sequences from different patients confirm the existence and transmissibility of subtype 2q, a rare subtype circulating in the metropolitan area of Barcelona, Spain.

Elisa Martro, Ana Valero, Elena Jordana-Lluch, Verónica Saludes, Ramón Planas, Fernando Gonzalez, Vicente Ausina, Maria Alma Bracho

► **To cite this version:**

Elisa Martro, Ana Valero, Elena Jordana-Lluch, Verónica Saludes, Ramón Planas, et al.. Hepatitis C virus sequences from different patients confirm the existence and transmissibility of subtype 2q, a rare subtype circulating in the metropolitan area of Barcelona, Spain.. *Journal of Medical Virology*, 2011, 83 (5), pp.820. 10.1002/jmv.22054 . hal-00624172

HAL Id: hal-00624172

<https://hal.science/hal-00624172>

Submitted on 16 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Hepatitis C virus sequences from different patients confirm the existence and transmissibility of subtype 2q, a rare subtype circulating in the metropolitan area of Barcelona, Spain.

Journal:	<i>Journal of Medical Virology</i>
Manuscript ID:	JMV-10-2119.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	24-Dec-2010
Complete List of Authors:	<p>Martro, Elisa; Hospital Universitari Germans Trias i Pujol, Microbiology Service Valero, Ana; Centre for Infectious Diseases. University of Edinburgh., Microbial Pathogenesis Unit. Jordana-Lluch, Elena; Hospital Universitari Germans Trias i Pujol, Microbiology Service Saludes, Verónica; Hospital Universitari Germans Trias i Pujol, Microbiology Service Planas, Ramón; Hospital Universitari Germans Trias i Pujol, Liver Unit Gonzalez, Fernando; University of Valencia, Instituto Cavanilles Biodiversidad y Biología Evolutiva Ausina, Vicente; Hospital Universitari Germans Trias i Pujol, Microbiology Service Bracho, Maria Alma; CSISP-UVEG/Institut Cavanilles, Joint Unit of Research in Genomics and Health</p>
Keywords:	HCV, genotype 2, subtype 2q, complete genome

SCHOLARONE™
Manuscripts

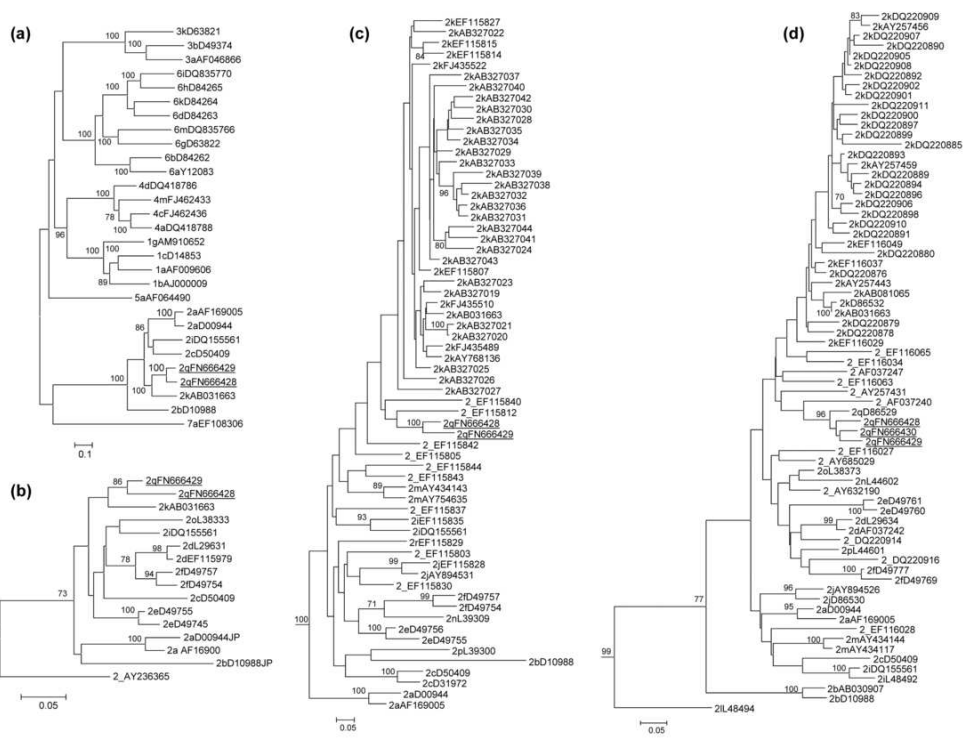


Figure 1. Maximum likelihood phylogenetic trees. (a) Unrooted phylogenetic tree of the coding region (positions 342 to 9336), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.301, gamma shape parameter: 0.841); (b) genotype 2 subtree of the Core region (positions 342 to 914), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.436, gamma shape parameter: 0.671); (c) genotype 2 subtree of the E1 region (positions 927 to 1262), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.436, gamma shape parameter: 0.671); (d) genotype 2 subtree of the NS5B region (positions 8283 to 8591), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.279, gamma shape parameter: 0.833). Positions are given according to the H77 reference sequence. Sequences are identified by genotype, subtype and GenBank accession number. Sequences that have not been assigned to any subtype are denoted as "2_". Sequences newly described in this study are underlined. Nodes supported by a bootstrap support value >70% are indicated. The scale bar represents number of substitutions per nucleotide position.

47x35mm (600 x 600 DPI)

1
2
3 **Hepatitis C virus sequences from different patients confirm the existence and**
4 **transmissibility of subtype 2q, a rare subtype circulating in the metropolitan area**
5 **of Barcelona, Spain.**
6
7
8
9

10 Elisa Martró,^{1,2*} Ana Valero,³ Elena Jordana-Lluch,¹ Verónica Saludes,^{1,2} Ramón
11 Planas,^{4,5} Fernando González-Candelas^{2,6}, Vicente Ausina^{1,7}, and Maria Alma Bracho.^{2,6}
12
13
14

15 ¹ Microbiology Service, Fundació Institut d'Investigació en Ciències de la Salut
16 Germans Trias i Pujol, Hospital Universitari Germans Trias i Pujol, Universitat
17 Autònoma de Barcelona, Badalona, Spain.
18
19
20
21

22 ² CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain.
23
24

25 ³ Microbial Pathogenesis Unit. Centre for Infectious Diseases. University of Edinburgh.
26
27 UK.
28

29 ⁴ Liver Unit, Hospital Universitari Germans Trias i Pujol, Badalona, Spain.
30
31

32 ⁵ CIBER Enfermedades Hepáticas y Digestivas (CIBEREHD), Barcelona, Spain.
33
34

35 ⁶ Joint Unit of Research in Genomics and Health CSISP-UVEG/Institut Cavanilles,
36
37 Valencia, Spain.
38

39 ⁷ CIBER Enfermedades Respiratorias (CIBERES), Bunyola, Spain.
40
41
42

43 **Shortened title:** Confirmation of HCV subtype 2q
44
45
46
47

48 *Corresponding author: Elisa Martró. Mailing address: Microbiology Service, Hospital
49 Universitari Germans Trias i Pujol, Ctra. del Canyet s/n, 08916 Badalona, Spain. Phone:
50
51 +34 934978 894. Fax: +34 934978 895. E-mail: emartro.igtp.germanstrias@gencat.cat
52
53
54
55
56
57
58
59
60

ABSTRACT

The hepatitis C virus (HCV) has been classified into 6 genotypes and more than 70 subtypes with distinct geographical and epidemiological distributions. While 18 genotype 2 subtypes have been proposed, only five have had their complete sequence determined. The aim of this study was to characterize HCV isolates from three patients from the Barcelona metropolitan area of Spain for whom commercial genotyping methods provided discordant results. Full-length genome sequencing was carried out for two of the three patients; for the third patient only partial NS5B sequences could be obtained. The generated sequences were subjected to phylogenetic, recombination and identity analyses. Sequences covering most of the HCV genome (9398 and 9566 nt in length) were obtained and showed a 90.3% identity to each other at the nucleotide level, while both sequences differed by 17.5-22.6% from the other fully sequenced genotype 2 subtypes. No evidence of recombination was found. The NS5B phylogenetic tree showed that sequences from the three patients cluster together with the only representative sequence of the provisionally designed 2q subtype, which also corresponds to a patient from Barcelona. Phylogenetic analysis of the full coding sequence showed that subtype 2q was more closely related to subtype 2k. The results obtained in this study suggest that subtype 2q now meets the requirements for confirmed designation status according to consensus criteria for HCV classification and nomenclature, and its epidemiological value is ensured as it has spread among several patients in the Barcelona metropolitan area.

KEY WORDS: HCV, genotype 2, subtype 2q, complete genome

INTRODUCTION

Hepatitis C virus (HCV) is estimated to infect 3% (170 million people) of the world's population, and is the most important cause of chronic liver disease and liver transplantation [World Health Organization 2010]. HCV is an enveloped virus belonging to the *Hepacivirus* genus within the *Flaviviridae* family. Its genome is a positive-sense single-stranded RNA of approximately 9.6 Kb. The HCV genome is flanked by noncoding (NC) regions, and includes a single open reading-frame encoding for both structural and nonstructural (NS) proteins [Choo et al. 1991]. HCV is characterized by a high level of genetic variability, which has allowed to classify the virus into 6 major genotypes (differing from each other by 31-33% at the nucleotide level), and a number of subtypes within each genotype (differing from each other by 20-25% at the nucleotide level) [Kuiken and Simmonds 2009]. On the basis of a single genomic sequence from a Canadian emigrant from the Democratic Republic of Congo, genotype 7a has been assigned provisionally [Murphy 2007]. Different genotypes show specific geographical locations, whilst certain subtypes (1a, 1b, 2a and 3a) have become widely distributed as a result of transmission through intravenous drug use, blood products, and health-care-related procedures [Kuiken and Simmonds 2009].

HCV genotype is a major predictor of a sustained virological response to antiviral therapy (pegylated alpha-interferon plus ribavirin); HCV genotyping is required to tailor dose and duration of treatment [Ghany et al. 2009]. The target of choice for commercialized molecular diagnostics and most genotyping assays is the 5'NC region, which is highly conserved. However, subtype identification based on this region is not always reliable. Although subtyping is not clinically relevant in the context of the current therapy, accurate subtype identification is useful for understanding the epidemiology and virological features of HCV. Besides, it might be clinically important

1
2
3 in the coming era of specifically targeted antiviral therapy for HCV, since subtype-
4 specific resistance profiles have been described for subtypes 1a and 1b both in infected
5 patients [Kieffer et al. 2007, Sarrazin et al. 2007] and *in vitro* [McCown et al. 2009].
6
7
8
9
10 Consensus criteria have been established for HCV classification and nomenclature, and
11 confirmed designation status of new subtypes with epidemiological value requires
12 rigorous phylogenetic analysis of one or more complete genome sequences of the
13 candidate subtype with no epidemiological connection [Kuiken and Simmonds 2009].
14
15
16
17
18
19
20 Eighteen subtypes within genotype 2 have been assigned so far (2a to 2r), but only four
21 have been confirmed (2a, 2b, 2c, and 2k) and a fifth one proposed (2i) on the basis of
22 their complete genome sequences.
23
24
25
26

27 In a previous study comparing three commercial kits for HCV-genotyping based
28 on the 5'NC region [Martró et al. 2008], three specimens that showed discordant results
29 were subjected to NS5B sequencing. Upon phylogenetic analysis, they did not cluster
30 with any of the previously confirmed genotype 2 subtypes, but formed a monophyletic
31 group with the only representative sequence of the provisionally assigned 2q subtype
32 [Tokita et al. 1998]. This study reports the HCV genomic sequence for two of those
33 isolates and a partial NS5B sequence for the third one, and suggests changing the status
34 of subtype 2q from proposed to confirmed designation.
35
36
37
38
39
40
41
42
43
44
45
46
47

48 METHODS

49
50
51 **Patients and specimens.** Three patients with a discordant genotyping result based on the
52 5'NC region were selected retrospectively from a previous study [Martró et al. 2008].
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1
2
3 Molecular Inc., Des Plaines, IL), while isolate 985 was genotyped as 2a/2c by reverse
4 hybridization, 2c by sequencing (TruGene HCV 5'NC genotyping kit, Bayer
5 HealthCare, Berkeley, CA), and 2a by the real-time PCR assay. Only by means of
6 NS5B sequencing and phylogenetic analysis evidence could be obtained showing that
7 the three isolates belonged probably to the previously proposed subtype 2q.
8

9
10 The clinical and epidemiological characteristics of the three patients are summarized in
11 Table I. Sera obtained at the date of HCV diagnosis had been conserved at -80°C.
12 Patient 985 had insufficient serum volume and only HCV isolates from patients 963 and
13 852 could be subjected to full-length genome sequencing.
14

15 This study was approved by the Clinical Research Ethics Committee ("Comité Ético de
16 Investigación Clínica", CEIC) at Hospital Universitari Germans Trias i Pujol
17 (HUGTIP).
18

19
20 **RNA extraction, RT-PCR and sequencing.** HCV RNA was extracted from 200 µl of
21 serum using the High Pure Viral RNA Kit (Roche) or QIAamp[®] viral RNA Mini kit
22 (QIAGEN[®]) following the manufacturer's protocol. Reverse transcription (RT) was
23 performed with 10 µl of extracted RNA in a total reaction volume of 20 µl containing 4
24 µl of 5× First-Strand buffer, 5 mM DTT, 200 U of SuperScript[®] III Reverse
25 Transcriptase (Invitrogen), 40 U of rRNasin[®] Ribonuclease Inhibitor (Promega), 0.5
26 mM of each dNTP, and either 0.75 µg of random hexadeoxynucleotides (Amersham-
27 Pharmacia Biotech) or 0.5 µM of antigenomic primer. The mix containing dNTP,
28 random hexadeoxynucleotides or primer, and the RNA was incubated at 65°C for 5 min.
29
30 Once the rest of components were added, the reaction was incubated at 50°C for 60 min
31 (for random hexadeoxynucleotides) or 55°C for 60 min (for primer), followed by 15 min
32 at 70°C.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Table I in supplementary material lists the primers used to obtain overlapping RT-PCR
4 products covering most of the viral genome. Primer design was initially based on
5 primers used in the sequencing of the 2i/6p recombinant (GenBank accession number
6 DQ155560) [Noppornpanth et al. 2006]. First round and hemi-nested PCR
7 amplifications were carried out in a final reaction volume of 50 μ l containing either 5 μ l
8 of the RT product (in the case of first round PCR) or 1 μ l of the first round PCR product
9 (in the case of hemi-nested PCR), 5 μ l of 10 \times PCR buffer with 15 mM MgCl₂, 200 μ M
10 of each dNTP, 400 nM each of both genomic sense and antigenomic sense primers, and
11 1 U of TaKaRa Taq™ (Takara Bio Europe SAS). Amplification was performed on 2700
12 or 9700 GeneAmp® PCR systems (Applied Biosystems) under the following conditions:
13 94°C for 2 min; 5 cycles, 94°C for 30 s, 50–65°C (depending on the primers used) for
14 30 s, 72°C for 3 min; 35 cycles, 94°C for 30 s, 52°C for 30 s and 72°C for 3 min; 72°C
15 for 10 min. For the 3'NC region, the following conditions were used: 94°C for 2 min; 5
16 cycles, 94°C for 30 s, 55°C for 1 min, 72°C for 2 min; 35 cycles, 94°C for 30 s, 52°C
17 for 1 min, 72°C for 2 min; 72°C for 10 min. Amplified products were checked by
18 electrophoresis on 1.8% agarose gels stained with ethidium bromide, and purified with
19 either High Pure PCR Products Purification Kit (Roche) or QIAquick PCR purification
20 kit (QIAGEN®).

21
22 Direct sequencing of PCR products was performed for both strands using the
23 BigDye™ Terminator v3.1 Ready Reaction Cycle Sequencing Kit on ABI Prism 3730
24 or 3100-Avant Genetic Analyzers (Applied Biosystems) and the primers listed in Table
25 II in supplementary material. The obtained chromatogram files were assembled,
26 verified, and edited using the Staden Package [Staden et al. 2000].
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The obtained sequences were deposited in the EMBL Nucleotide Sequence Database
4 with accession numbers FN666428 (patient 963), FN666429 (patient 852), and
5
6
7
8
9 FN666430 (patient 985).

10 *Phylogenetic reconstructions and identity analysis.* The nucleotide sequences
11
12 corresponding to the complete polyprotein from patients 963 and 852 were aligned with
13
14
15 27 homologous sequences representative of the main HCV genotypes and subtypes (see
16
17 GenBank accession numbers, genotypes and subtypes in Figure 1) with ClustalW
18
19 [Thompson et al. 1994] implemented in MEGA v4.0 [Tamura et al. 2007]. Alignments
20
21 corresponding to partial genomic regions (Core, E1 and NS5B) were also obtained. The
22
23 alignments included the most similar sequences found in a BLAST search against HCV
24
25 sequences, including all 2k deposited sequences and sequences for all genotype 2
26
27 subtypes available, as well as reference sequences for the other genotypes. The
28
29 sequence from patient 985 was also considered for the analysis of the NS5B region.
30
31
32

33
34 Maximum likelihood phylogenetic trees were constructed from the corresponding
35
36 alignments with the evolutionary model implemented in PHYML that best fitted the
37
38 data according to jModeltest [Posada 2008]. The robustness of the tree topologies was
39
40 assessed by bootstrap analysis implemented in PHYML [Guindon and Gascuel 2003]
41
42 with 1000 pseudorandom resamplings of the sequences. GenBank accession numbers of
43
44 the reference sequences used in each phylogenetic reconstruction are shown in Figure 1.
45
46
47

48 Pairwise nucleotide identities between genotype 2 subtypes including the newly
49
50 described 2q sequences were calculated using MEGA v4.0, using an alignment
51
52 encompassing nucleotide positions 31 (5'NC region) to 9336 (end of the 2i sequence
53
54 DQ155561) of the H77 reference sequence (GenBank accession number AF009606).
55
56
57 Reference sequences for the following HCV subtypes were used in this analysis:
58
59
60

1
2
3 Subtype 2a (GenBank accession number D00944), 2b (D10988), 2c (D50409), 2k
4
5 (AB031663) and 2i (DQ155561).
6
7

8 **Recombination analysis.** Recombination analysis was performed using the
9
10 recombination detection program RDP v3.44 [Martin and Rybicki 2000] with an
11
12 alignment that included the two genomic sequences generated in this study along with
13
14 the 27 representative complete sequences used in the phylogenetic analyses [Heath et al.
15
16 2006]. This program implements several methods for the identification of recombinant
17
18 sequences and recombination breakpoints. Recombination analysis was performed using
19
20 six of them: two phylogenetic methods that detect recombination when different regions
21
22 of the genome lead to discordant topologies (RDP, and Bootscan/Recscan [Martin et al.
23
24 2005]), and four nucleotide substitution methods that search either for a significant
25
26 clustering of substitutions or for a fit to an expected statistical distribution (GeneConv
27
28 [Padidam et al. 1999], Chimaera [Posada and Crandall 2001], MaxChi [Maynard Smith
29
30 1992], and SiScan [Gibbs et al. 2000]). Default RDP v3.44 settings were used
31
32 throughout, except that sequences were considered to be linear. Only putative
33
34 recombination events inferred by more than two methods were taken into account, as
35
36 performance evaluation of these recombination detection methods indicated that the
37
38 results of a single method are not completely reliable [Posada 2002].
39
40
41
42
43
44
45
46
47

48 **RESULTS**

49
50 The HCV genomic sequences obtained were 9398 and 9566 nt in length for
51
52 patients 852 and 963, respectively. Both sequences were partial at the 5' end of the
53
54 5'NC region; the first one ended 18 nt before the end of the polyprotein coding region
55
56 (the NS5B region is partial at the 3' end), and the later included part of the 3'NC region.
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1a shows the unrooted ML phylogenetic tree corresponding to the polyprotein-encoding nucleotide sequence. For genotype 2, only the five fully sequenced subtypes could be included in the analysis (2a, 2b, 2c, 2i and 2k), since complete genome sequences are still lacking for the other 13 assigned subtypes. This tree shows a highly supported grouping of sequences from patients 852 and 963, which would constitute a highly supported clade along with subtype 2k. Similar results were obtained in the phylogenetic analysis of the full Core region (573 nt) in which a few additional subtype 2 sequences were included (Figure 1b).

Phylogenetic analysis of partial E1 and NS5B regions provided additional information. The phylogenetic tree of a NS5B subregion (309 nt) included the three patients considered in this study (852, 963 and 985, for which only this partial NS5B sequence could be obtained). As shown in Figure 1d, these three sequences cluster together with the only representative sequence of the provisionally assigned 2q subtype (GenBank accession number D86529) with a high bootstrap support. Interestingly, the latter corresponds to isolate BA045, which was obtained in the 90's from a patient residing in Barcelona, Spain, who had not travelled abroad, and had acquired HCV infection through transfusion (M. A. Barrera, personal communication). Sequence AF037240 from a patient from Burkina Faso (West Africa) and described as a putative new genotype 2 subtype [Jeannel et al. 1998], was the sequence related most closely to the 2q cluster. Similarly, the phylogenetic tree of the E1 region (336 nt) shows that the two 2q sequences clustered with maximum bootstrap support (Figure 1c), and nontypable sequences EF115812 and EF115840 from patients in Quebec, Canada [Murphy et al. 2007], were the most closely related sequences.

Recombination signals were investigated by means of phylogenetic and nucleotide substitution methods. These analyses based on the complete genome alignment showed

1
2
3 no evidence that the two newly described 2q genomes had been involved in
4
5 recombination events involving other genotypes or genotype 2 subtypes.
6
7

8 Finally, when considering the coding region, the two subtype 2q genomes
9
10 sequenced showed a 90% identity to each other at the nucleotide level, while both
11
12 sequences differed by 17.5 to 22.6% from the other confirmed genotype 2 subtypes
13
14 (Table II). The highest identity was observed between subtypes 2q and 2c (82.23%)
15
16 while the identity between 2q and 2k was slightly lower (80.54%). However, genetic
17
18 distances estimated by maximum likelihood (which attempt to estimate the true amount
19
20 of change occurred between two subtypes) were 0.40 and 0.28 respectively, which
21
22 reveals the closest relationship between 2q and 2k subtypes. When the length of the full
23
24 coding region as well as the ten protein regions were compared between genotype 2
25
26 subtypes, all subtype 2q protein regions were identical in length to those of subtypes 2a,
27
28 2b and 2k (Table III).
29
30
31
32
33
34
35

36 DISCUSSION

37
38 Genetic diversity is greatest within HCV genotypes 6, 3 and 2, which contains
39
40 more divergent sequences than other genotypes [Kuiken and Simmonds 2009].
41
42 Genotype 2 is highly prevalent and genetically diverse in sub-Saharan Africa and in
43
44 Martinique [Candotti et al. 2003, Jeannel et al. 1998, Martial et al. 2004, Ruggieri et al.
45
46 1996]. In Europe, previous studies found a prevalence of 11% in France, with a
47
48 considerable diversity [Thomas et al. 2007], and a 3% in Spain [Echevarria et al. 2006].
49
50 Genotype 2 represents 4.3% of the patients treated in the Hospital Universitari Germans
51
52 Trias i Pujol (period 2003-2009, N=493), where a variety of subtypes have been
53
54 identified, including 2a, 2b, 2c, 2i, 2j and 2q [Martró et al. 2008].
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

New genotype 2 subtypes with limited distribution have been proposed on the basis of a few partial genome sequences in certain geographical regions not only in Africa [Jeannel et al. 1998, Ruggieri et al. 1996], but also in Europe [Thomas et al. 2007, van Doorn et al. 1995]. Among HCV-2 subtypes, 2a, 2b, 2c and 2k are worldwide distributed while subtypes 2d to 2p, 2r and five other tentative subtypes [Cantaloube et al. 2008] have been detected in only one or a few countries. HCV 2q, which had been proposed previously on the basis of a single partial NS5B sequence from a patient from Barcelona [Tokita et al. 1998] is another example of a geographically restricted subtype. In this study, another three patients in the Barcelona metropolitan area were identified as bearing this subtype, and most of its genome has been sequenced in two of the three cases. Patients' ages and clinical records agree with the hypothesis that HCV-2q could have been emerged in Spain as a result of blood transfusion, probably before HCV screening was implemented in blood banks in 1990. In fact, the most common transmission route among patients infected by minority genotype 2 subtypes in the close area of Midi-Pyrenées in France was through blood transfusion, while subtype 2a was transmitted via intravenous drug use [Thomas et al. 2007].

Three commercial genotyping assays based on the 5'NC region either failed to assign a genotype or to subtype correctly the three 2q isolates. Nucleotide sequencing and phylogenetic analysis of more variable genomic regions has been recommended for HCV genotyping in consensus proposals [Simmonds et al. 2005]. Accordingly, commercialized assays based on the Core (Versant HCV genotype 2.0 assay, Siemens Medical Solutions Diagnostics) or the NS5B regions (Abbott RealTime HCV Genotype II, Abbott Molecular Diagnostics) have been developed over the last years for a better discrimination of genotype 1 subtypes. However, subtype assignment for the other genotypes still relies on the 5'NC region and often leads to errors [Chen and Weck

1
2
3 2002, Martró et al. 2008, Tamalet et al. 2003, Zeuzem et al. 1995]. Furthermore, only
4
5 sequencing and phylogenetic analysis of coding regions, such as NS5B or Core,
6
7 provides the chance to identify variants not yet classified.
8
9

10 A level of nucleotide differences of at least 15% is expected between subtypes
11
12 within each genotype, including newly described variants [Kuiken and Simmonds
13
14 2009]. Accordingly, subtype 2q shows a nucleotide difference of 17.5-22.6% from the
15
16 other confirmed genotype 2 subtypes. Phylogenetic analysis of the coding region
17
18 showed that, among confirmed genotype 2 subtypes, subtype 2q was most closely
19
20 related to subtype 2k. Furthermore, in the phylogenetic analysis of the NS5B and E1
21
22 regions, non-typable genotype 2 sequences in patients from Burkina Faso and Canada
23
24 were identified as the most closely related sequences to subtype 2q. Several
25
26 recombinants of different genotypes including genotype 2 have been described in
27
28 several countries: 2k/1b [Kalinina et al. 2004], 2i/6p [Noppornpanth et al. 2006], and
29
30 2b/1b [Kageyama et al. 2006], and 2k/5 [Legrand-Abravanel et al. 2007]. Nevertheless,
31
32 the two newly described 2q sequences showed no evidence of recombination.
33
34
35
36
37

38 In conclusion, phylogenetic, identity, and recombination analyses indicate that
39
40 subtype 2q now meets the requirements for confirmed designation status according to
41
42 consensus criteria for HCV classification and nomenclature [Kuiken and Simmonds
43
44 2009, Simmonds et al. 2005]. Furthermore, subtype 2q is of epidemiological value,
45
46 since it has spread among several patients in the Barcelona metropolitan area.
47
48
49
50
51

52 **ACKNOWLEDGEMENTS**

53
54 This work was supported partially by CIBER Epidemiología y Salud Pública
55
56 (CIBERESP), projects PI05/1131 (EM, RP, VA), CP09/0004 (EM), CP07/00078
57
58 (MAB), and BFU2008-03000 (FG-C, MAB) from “Instituto de Salud Carlos III - Fondo
59
60

1
2
3 de Investigaciones Sanitarias” (ISCIH-FIS), and grants CD05/00258 (“Contratos
4 Postdoctorales de Perfeccionamiento”) and “Miguel Servet” from “Ministerio de
5
6 Ciencia e Innovacion” (MICINN), within the “Plan Nacional de Investigación científica,
7
8 Desarrollo e Innovación Tecnológica (I+D+I)” (EM).
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

REFERENCES

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Candotti D, Temple J, Sarkodie F, Allain JP. 2003. Frequent recovery and broad genotype 2 diversity characterize hepatitis C virus infection in Ghana, West Africa. *J Virol* 77:7914-7923.

Cantaloube JF, Gallian P, Laperche S, Elghouzzi MH, Piquet Y, Bouchardeau F, Jordier F, Biagini P, Attoui H, de Micco P. 2008. Molecular characterization of genotype 2 and 4 hepatitis C virus isolates in French blood donors. *J Med Virol* 80:1732-1739.

Chen Z, Weck KE. 2002. Hepatitis C virus genotyping: interrogation of the 5' untranslated region cannot accurately distinguish genotypes 1a and 1b. *J Clin Microbiol* 40:3127-3134.

Choo QL, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos D, Coit A, Medina-Selby P, Barr J, Weiner AJ, Bradley DW, Kuo G, and Houghton M. 1991. Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci USA* 88:2451-2455.

Echevarria JM, Leon P, Pozo F, Avellon A. Follow-up of the prevalence of hepatitis C virus genotypes in Spain during a nine-year period (1996-2004). 2006. *Enferm Infecc Microbiol Clin* 24:20-25.

Ghany MG, Strader DB, Thomas DL, Seeff LB. 2009. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* 49:1335-1374.

Gibbs MJ, Armstrong JS, Gibbs AJ. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16:573-582.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696-704.

Heath L, van der WE, Varsani A, Martin DP. 2006. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J Virol* 80:11827-11832.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

eannel D, Fretz C, Traore Y, Kohdjo N, Bigot A, Pe GE, Jourdan G, Kourouma K, Maertens G, Fumoux F, Fournel JJ, Stuyver L. 1998. Evidence for high genetic diversity and long-term endemicity of hepatitis C virus genotypes 1 and 2 in West Africa. *J Med Virol* 55:92-97.

Kageyama S, Agdamag DM, Alesna ET, Leano PS, Heredia AM, Abellanos-Tac-An IP Jereza LD, Tanimoto T, Yamamura J, Ichimura H. 2006. A natural inter-genotypic (2b/1b) recombinant of hepatitis C virus in the Philippines. *J Med Virol* 78:1423-1428.

Kalinina O, Norder H, Magnus LO. 2004. Full-length open reading frame of a recombinant hepatitis C virus strain from St. Petersburg: proposed mechanism for its formation. *J Gen Virol* 85:1853-1857.

Kieffer TL, Sarrazin C, Miller JS, Welker MW, Forestier N, Reesink HW, Kwong AD, Zeuzem S. 2007. Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients. *Hepatology* 46:631-639.

Kuiken C, Simmonds P. 2009. Nomenclature and numbering of the hepatitis C virus. In: Tang H, editor. *Hepatitis C: Methods and Protocols*. Humana Press, p 33-53.

Legrand-Abravanel F, Claudinon J, Nicot F, Dubois M, Chapuy-Regaud S, Sandres-Saune K, Pasquier C, Izopet J. 2007. New natural intergenotypic (2/5) recombinant of hepatitis C virus. *J Virol* 81:4357-4362.

Martial J, Morice Y, Abel S, Cabié A, Rat C, Lombard F, Edouard A, Pierre-Louis S, Garsaud P, Béra O, Chout R, Gordien E, Deny P, Césaire R. 2004. Hepatitis C virus (HCV) genotypes in the Caribbean island of Martinique: evidence for a large radiation of HCV-2 and for a recent introduction from Europe of HCV-4. *J Clin Microbiol* 42:784-791.

Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16:562-563.

1
2
3 Martin DP, Posada D, Crandall KA, Williamson C. 2005. A modified bootscan
4 algorithm for automated identification of recombinant sequences and recombination
5 breakpoints. *AIDS Res Hum Retroviruses* 21:98-102.
6
7

8
9
10 Martró E, González V, Buckton AJ, Saludes V, Fernández G, Matas L, Planas R,
11 Ausina V. 2008. Evaluation of a new assay in comparison with reverse hybridization
12 and sequencing methods for hepatitis C virus genotyping targeting both 5' noncoding
13 and nonstructural 5b genomic regions. *J Clin Microbiol* 46:192-197.
14
15

16
17
18 Maynard Smith J. 1992. Analyzing the mosaic structure of genes. *J Mol Evol* 34:126-
19 129.
20
21

22
23
24 McCown MF, Rajyaguru S, Kular S, Cammack N, Najera I. 2009. GT-1a or GT-1b
25 subtype-specific resistance profiles for hepatitis C virus inhibitors telaprevir and HCV-
26 796. *Antimicrob Agents Chemother* 53:2129-2132.
27
28

29
30
31 Murphy DG, Willems B, Deschenes M, Hilzenrat N, Mousseau R, Sabbah S. 2007. Use
32 of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus
33 with reference to C/E1 and 5' untranslated region sequences
34 10. *J Clin Microbiol* 45:1102-1112.
35
36

37
38
39 Noppornpanth S, Lien TX, Poovorawan Y, Smits SL, Osterhaus AD, Haagmans BL.
40 2006. Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus.
41 *J Virol* 80:7569-7577.
42
43

44
45
46 Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses
47 by frequent recombination. *Virology* 265:218-225.
48
49

50
51
52 Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25:1253-
53 1256.
54
55

56
57
58 Posada D. 2002. Evaluation of methods for detecting recombination from DNA
59 sequences: empirical data. *Mol Biol Evol* 19:708-717.
60

- 1
2
3 Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from
4
5 DNA sequences: computer simulations. *Proc Natl Acad Sci USA* 98(24):13757-13762.
6
7
8 Ruggieri A, Argentini C, Kouruma F, Chionne P, D'Ugo E, Spada E, Dettori S,
9
10 Sabbatani S, Rapicetta M. 1996. Heterogeneity of hepatitis C virus genotype 2 variants
11
12 in West Central Africa (Guinea Conakry). *J Gen Virol* 77:2073-2076.
13
14
15 Sarrazin C, Kieffer TL, Bartels D, Hanzelka B, Muh U, Welker M, Wincheringer D,
16
17 Zhou Y, Chu HM, Lin C, Weegink C, Reesink H, Zeuzem S, Kwong AD. 2007.
18
19 Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the
20
21 protease inhibitor telaprevir. *Gastroenterology* 132:1767-1777.
22
23
24 Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, Halfon P,
25
26 Inchauspé G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H,
27
28 Pawlotsky JM, Penin F, Sablon E, Shin-I T, Stuyver LJ, Thiel HJ, Viazov S, Weiner AJ,
29
30 Widell A. 2005. Consensus proposals for a unified system of nomenclature of hepatitis
31
32 C virus genotypes. *Hepatology* 42:962-973.
33
34
35 Staden R, Beal KF, Bonfield JK. The Staden package, 1998. 2000. *Methods Mol Biol*
36
37 132:115-130.
38
39
40
41 Tamalet C, Colson P, Tissot-Dupont H, Henry M, Tourres C, Tivoli N, Botta D, Ravaux
42
43 I, Poizot-Martin I, Yahi N. 2003. Genomic and phylogenetic analysis of hepatitis C
44
45 virus isolates: a survey of 535 strains circulating in southern France. *J Med Virol*
46
47 71:391-398.
48
49
50
51 Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary
52
53 Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599.
54
55
56 Thomas F, Nicot F, Sandres-Saune K, Dubois M, Legrand-Abravanel F, Alric L, Peron
57
58 JM, Pasquier C, Izopet J. 2007. Genetic diversity of HCV genotype 2 strains in south
59
60 western France. *J Med Virol* 79:26-34.

1
2
3 Thomas F, Nicot F, Sandres-Saune K, Dubois M, Legrand-Abravanel F, Alric L Peron
4 JM, Pasquier C, Izopet J. 2007. Genetic diversity of HCV genotype 2 strains in south
5 western France. *J Med Virol* 79:26-34.
6
7

8
9
10 Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity
11 of progressive multiple sequence alignment through sequence weighting, position-
12 specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680.
13
14

15
16
17 Tokita H, Okamoto H, Iizuka H, Kishimoto J, Tsuda F, Miyakawa Y, Mayumi M. 1998.
18 The entire nucleotide sequences of three hepatitis C virus isolates in genetic groups 7-9
19 and comparison with those in the other eight genetic groups. *J Gen Virol* 79:1847-1857.
20
21

22
23
24 van Doorn LJ, Kleter GE, Stuyver L, Maertens G, Brouwer JT, Schalm SW, Heijtkink
25 RA, Quint WG. 1995. Sequence analysis of hepatitis C virus genotypes 1 to 5 reveals
26 multiple novel subtypes in the Benelux countries. *J Gen Virol* 76:1871-1876.
27
28

29
30
31 World Health Organization. 2010. Viral Cancers: Hepatitis C virus.
32 [www.who.int/vaccine_research/diseases/viral_cancers/en/index2.html#disease%20burd](http://www.who.int/vaccine_research/diseases/viral_cancers/en/index2.html#disease%20burden)
33 [en](http://www.who.int/vaccine_research/diseases/viral_cancers/en/index2.html#disease%20burden).
34
35

36
37
38 Zeuzem S, Ruster B, Lee JH, Stripf T, Roth WK. 1995. Evaluation of a reverse
39 hybridization assay for genotyping of hepatitis C virus. *J Hepatol* 23:654-661.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FIGURE LEGENDS

Figure 1. Maximum likelihood phylogenetic trees. (a) Unrooted phylogenetic tree of the coding region (positions 342 to 9336), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.301, gamma shape parameter: 0.841); (b) genotype 2 subtree of the Core region (positions 342 to 914), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.436, gamma shape parameter: 0.671); (c) genotype 2 subtree of the E1 region (positions 927 to 1262), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.436, gamma shape parameter: 0.671); (d) genotype 2 subtree of the NS5B region (positions 8283 to 8591), nucleotide substitution model GTR + I + G (proportion of invariable sites: 0.279, gamma shape parameter: 0.833). Positions are given according to the H77 reference sequence. Sequences are identified by genotype, subtype and GenBank accession number. Sequences that have not been assigned to any subtype are denoted as “2_”. Sequences newly described in this study are underlined. Nodes supported by a bootstrap support value >70% are indicated. The scale bar represents number of substitutions per nucleotide position.

TABLES

Table I. Characteristics of the four patients identified so far as being infected with HCV subtype 2q.

Characteristic	Patient ID			
	852	963	985	BA045
Gender	Male	Female	Female	Male
HCV risk factor	Surgery, several transfusions (1959)	Multiparous, two abortions (<1990), curettage (1992)	Cervical carcinoma surgery (1990), possible transfusion	Transfusion
Date of HCV diagnosis	2001	2002	2001	-
Age (years)	60	51	69	-
HCV treatment	No	No	Not eligible for treatment	Yes
Clinical outcome	Persistently normal transaminase levels Discharged (2005)	Cirrhosis, portal hypertension (esophageic varices)	Pluripathology (advanced renal insufficiency) Exitus (2005)	Sustained virological response
City of residence (province)	Santa Coloma (Barcelona)	Badalona (Barcelona)	Badalona (Barcelona)	Barcelona (Barcelona)

Table II. Identity (%) among genotype 2 subtypes that have been characterized by complete genome sequencing. GenBank accession numbers of reference sequences: D00944 (2a), D10988 (2b), D50409 (2c), AB031663 (2k) and DQ155561 (2i). Nucleotide identity was computed between nucleotide positions 31 to 9336 according to the H77 reference sequence. Patient ID is indicated in parenthesis.

	2a	2b	2c	2i	2k	2q (852)
2a	-					
2b	76.75	-				
2c	79.44	76.09	-			
2i	79.93	77.33	80.29	-		
2k	80.16	76.64	80.09	81.10	-	
2q (852)	79.62	77.42	82.52	80.40	79.47	-
2q (963)	80.04	77.58	82.23	80.54	80.10	90.34

Table III. Size composition of HCV protein regions among the genotype 2 subtypes that have been characterized by complete genome sequencing. ^aBased on curated alignments available at the HCV LANL Database (<http://hcv.lanl.gov/content/sequence/HCV/ToolsOutline.html>). ^bPartial sequence at its 3' end. ^cBased only on the 963 isolate. GenBank accession numbers of reference sequences: D00944 (2a), D10988 (2b), D50409 (2c), AB031663 (2k) and DQ155561 (2i).

	Region length (nt) ^a										
	Core	E1	E2	p7	NS2	NS3	NS4A	NS4B	NS5A	NS5B	Polyprotein
2a	573	576	1101	189	651	1893	162	783	1398	1776	9102
2b	573	576	1101	189	651	1893	162	783	1398	1776	9102
2c	573	576	1101	189	651	1893	162	783	1410	1776	9114
2i	573	576	1101	189	651	1893	162	783	1395	1734 ^b	9057 ^b
2k	573	576	1101	189	651	1893	162	783	1398	1776	9102
2q	573	576	1101	189	651	1893	162	783	1398	1776 ^c	9102 ^c