

Supplementary Table S1. *In silico* analysis of pneumococcal serotypes that can be amplified and differentiated by sequencing based on sequence diversity of *cpsB*.

Serotype (n = 92)^b	Amplification possible^{a,c}	Differentiable by <i>in silico</i> analysis^{a, c, d}	Differentiable by <i>cpsB</i> PCR and sequencing?^{a, c, d}	Included in <i>in vitro</i> testing^a
1	Y	Y	Y	Y
2	Y	(2/41A)	(2/41A)	N
3	Y	Y	Y	Y
4	Y	Y	Y	Y
5	Y	Y	Y	Y
6A	Y	Y	Y	Y
6B	Y	Y	Y	Y
6C	Y	Y	(6C/6D)	Y
6D	Y	Y	(6C/6D)	N
7F	Y	Y	(7A/7F)	Y
7A	Y	Y	(7A/7F)	N
7B	Y	(7B/40)	(7B/40)	N
7C	Y	Y	Y	Y
8	Y	Y	Y	Y
9A	Y	(9A/9V)	(9A/9V)	N
9L	Y	Y	Y	N
9N	Y	Y	Y	Y
9V	Y	(9V/9A)	(9V/9A)	Y
10F	Y	(10F/10C)	(10F/10C)	Y
10A	Y	Y	Y	Y
10B	Y	Y	Y	N
10C	Y	(10C/10F)	(10C/10F)	N
11F	N	Unknown	Unknown	N
11A	Y	(11A/11D/18F)	(11A/11D/18F)	Y
11B	Y	(11B/11C)	(11B/11C)	N
11C	Y	(11B/11C)	(11B/11C)	N
11D	Y	(11A/11D/18F)	(11A/11D/18F)	N
12F	Y	(12F/44)	(12F/44)	Y
12A	Y	Y	Y	N
12B	Y	Y	Y	Y
13	Y	(13/20)	(13/20)	N
14	Y	Y	Y	Y
15F	Y	Y	Y	N
15A	Y	Y	Y	Y
15B	Y	Y	Y	Y
15C	Y	Y	Y	N
16F	Y	Y	Y	Y
16A	Y	Y	Y	N
17F	Y	Y	(17F/33C)	Y
17A	Y	(17A/34)	(17A/34)	Y
18F	Y	(11A/11D/18F)	(11A/11D/18F)	N
18A	Y	Y	Y	N
18B	Y	Y	(18B/18C)	Y
18C	Y	Y	(18B/18C)	Y

19F	Y	Y	Y	Y
19A	Y	Y	Y	Y
19B	Y	Y	Y	N
19C	Y	Y	Y	N
20	Y	(20/13)	(20/13)	Y
21	Y	Y	Y	Y
22F	Y	(22F/22A)	(22F/22A)	Y
22A	Y	(22A/22F)	(22A/22F)	Y
23F	Y	Y	Y	Y
23A	Y	Y	Y	N
23B	Y	Y	Y	Y
24F	Y	Y	Y	Y
24A	Y	Y	Y	N
24B	Y	Y	Y	Y
25F	N	Unknown	Unknown	N
25A	N	Unknown	Unknown	N
27	N	Unknown	Unknown	Y
28F	Y	Y	Y	Y
28A	Y	Y	Y	Y
29	N	Unknown	Unknown	Y
31	Y	Y	Y	Y
32F	Y	(32F/32A)	(32F/32A)	N
32A	Y	(32A/32F)	(32A/32F)	N
33F	Y	(33F/33A/35A)	(33F/33A/35A)	Y
33A	Y	(33A/33F/35A)	(33A/33F/35A)	N
33B	Y	Y	Y	Y
33C	Y	Y	(17F/33C)	Y
33D	Y	Y	Y	Y
34	Y	(34/17A)	(34/17A)	Y
35F	Y	(35F/47F)	(35F/47F)	N
35A	Y	(35A/33F/33A)	(35A/33F/33A)	Y
35B	Y	(35B/35C)	(35B/35C)	Y
35C	Y	(35C/35B)	(35C/35B)	N
36	Y	Y	Y	Y
37	Y	Y	Y	N
38	N	Unknown	Unknown	N
39	N	Unknown	Unknown	N
40	Y	(7B/40)	(7B/40)	N
41F	Y	Y	Y	N
41A	Y	(2/41A)	(2/41A)	N
42	Y	Y	Y	N
43	N	Unknown	Unknown	N
44	Y	(12F/44)	(12F/44)	N
45	Y	Y	Y	N
46	Y	Y	Y	N
47F	Y	(35F/47)	(35F/47)	N
47A	Y	Y	Y	N
48	Y	Y	Y	N
Total	Y = 84/92 N = 8/92	Y=54/92 Ambiguous =	Y = 46/92 Ambiguous =	Y = 48/92 N = 44

30/92 **38/92**
Unknown=8/92 **Unknown = 8/92**

^a Y = yes, N = no.

^b Serotypes of PPV23 and PCV13 are indicated in bold.

^c Serotypes that are amplifiable by PCR but give ambiguous results are indicated as two or more serotypes having the same top BLAST bit scores.

^d Serotypes deemed non-amplifiable in initial *in silico* analysis are not included in the subsequent tests on identifiable serotypes, therefore resolution power is unknown for these serotypes.