

```
/* -----  
Sequencing Plate Automatic Renaming Utility (SPARE).  
C R Young  
last updated 4/20/08
```

Copyright (C) 2008 C R Young

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <<http://www.gnu.org/licenses/>>.

LEGAL DISCLAIMER

By using this code you agree that I am in no way responsible for any damages that may befall you due to my incompetence, negligence, or any other nasty thing that you accuse me of. After all, you could do this by hand.

```
-----*/
```

This program renames sequences input in a fasta file format based on a name lookup table supplied in a tab delimited text file. The output is also a fasta formatted file. Usage is:

```
./SPARE [fasta data file] [name lookup table] [sample number position]
```

where the options are:

[fasta data file]	= input file to be renamed
[name lookup table]	= tab delimited name lookup table
[sample number position]	= character position in the sequence name that denotes plate position

So that to rename the example files, one would type:

```
./SPARE input_seqs.fas lookup.txt 10
```

The program assumes that the names of the sequences that are supplied as input contain, somewhere in the sequence name, a number (1–96) that corresponds to the plate position. For example, the following sequence is position number one on the plate:

```
>Contig[0001]
TAGTCCGGATCGGAG...
```

The user needs to tell the program that the plate number starts at position 10 in the name (Note: This position must be the same for all sequences! and the “>” character should not be counted!)

To create the sample lookup table, I use excel. A sequencing plate might look like this:

Table 1. Sample names as supplied in the lookup table

	1	2	3	4	5	6	7	8	9	10	11	12
A	233-1-1	233-1-9	233-1-17	233-2-1	233-2-9	233-2-17	240-1-1	240-1-9	240-1-17	240-2-1	240-2-9	240-2-17
B	233-1-2	233-1-10	233-1-18	233-2-2	233-2-10	233-2-18	240-1-2	240-1-10	240-1-18	240-2-2	240-2-10	240-2-18
C	233-1-3	233-1-11	233-1-19	233-2-3	233-2-11	233-2-19	240-1-3	240-1-11	240-1-19	240-2-3	240-2-11	240-2-19
D	233-1-4	233-1-12	233-1-20	233-2-4	233-2-12	233-2-20	240-1-4	240-1-12	240-1-20	240-2-4	240-2-12	240-2-20
E	233-1-5	233-1-13	233-1-21	233-2-5	233-2-13	233-2-21	240-1-5	240-1-13	240-1-21	240-2-5	240-2-13	240-2-21
F	233-1-6	233-1-14	233-1-22	233-2-6	233-2-14	233-2-22	240-1-6	240-1-14	240-1-22	240-2-6	240-2-14	240-2-22
G	233-1-7	233-1-15	233-1-23	233-2-7	233-2-15	233-2-23	240-1-7	240-1-15	240-1-23	240-2-7	240-2-15	240-2-23
H	233-1-8	233-1-16	233-1-24	233-2-8	233-2-16	233-2-24	240-1-8	240-1-16	240-1-24	240-2-8	240-2-16	240-2-24

Where positions on the plate are assumed to be numbered horizontally as such:

Table 2. Plate position corresponding to the sequence names in the input file

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	2	3	4	5	6	7	8	9	10	11	12
B	13	14	15	16	17	18	19	20	21	22	23	24
C	25	26	27	28	29	30	31	32	33	34	35	36
D	37	38	39	40	41	42	43	44	45	46	47	48
E	49	50	51	52	53	54	55	56	57	58	59	60
F	61	62	63	64	65	66	67	68	69	70	71	72
G	73	74	75	76	77	78	79	80	81	82	83	84
H	85	86	87	88	89	90	91	92	93	94	95	96

The lookup table supplied to the program should not contain the column headers (1–12) or row headers (A–H). See the example file supplied with the program. The excel file corresponding to Table 1 can be saved directly as a tab-delimited text file. (Note: If one needs to rename a plate that has been numbered vertically, you just need to transpose the matrix in excel and save that as the lookup table.) Note that the input sequence file need not contain 96 sequences (e.g. if some sequences on the plate didn’t work), but the lookup table must include the names of all 96 sequences on the plate.

The renamed sequences are output in the file Seq_data.fas.

To compile the program using a linux environment, type:

```
gcc SPARE.c -o SPARE
```