SVARAP and aSVARAP: simple tools for quantitative analysis of nucleotide.

SVARAP user manual

SVARAP and aSVARAP are free and downloadable at URL: http://ifr48.timone.univmrs.fr/portail2/index.php?option=com_content&task=view&id=50#telechargements

-To get an alignment in GDE format using clustal X v1.83

Clustal X v1.83 is free and downloadable at URL: ftp-igbmc.u-strasbg.fr/pub/ClustalX/

- Open ClustalX (1.8) and append sequences in FASTA format.
- Select tab « Alignment », then output Format Options
- Select GDE format.
- Start alignment ("Do complete alignment")



ClustalX (1.8) File Edit Alignment Trees Colors	Quality Help		
Multiple Alignment Mode ▼ 1 Pr0046458 2 Pr0046459 3 Pr0046450 4 Pr0046451 5 Pr0046451 5 Pr0046453 7 Pr0046453 7 Pr0046454 8 Pr0046459 9 Pr0046588 10 Pr0046588 10 Pr0046588	Font Size: 10 Output Format Options CLOSE Output Files CLUSTAL format NBRF/P GCG/MSF format PHYLIP GDE format	IR format	A CACATIGA GGGTCA GCAGTA ACACATIGA RGGTCA GCAGTA ACATATIGA RGGTCA GCAGTA ACATATIGA RGGTCAGCA GTA ACATATIGA RGGTCAGCA GTA ACATATIGA RGGTCAGCA GTA ACATATIGA RGGTCAGCAGTA ATACATIGA GGGTCAGCAGTA ATACATIGA GGGTCAGCAGTA ATACATIGA GGGTCAGCAGTA
11 Pr0046891 12 Pr0046893 13 Pr0046893 14 Pr0046919 15 Pr0046920 16 Pr0046921 17 Pr0046922 18 Pr0046923 19 Pr0046923	GDE output case : CLUSTALW sequence numbers Output order	Lower COFF LOFF LIGNED	A TACATTGA GGGTCA GCCA GTA ATATATCCA GGGTCA GCCA GTA A TACATTGA GGGTCA GCCA GTA A TACATTGA GGCTCA GCCA GTA A TACATTGA GGGTCA GCCA GTA A TACATTCA GGGTCA GCCA GTA A CACATTCA GGGTCA GCCA GTA A CACATTCA GGGTCA GCCA GTA
File C:\SeqProt.txt appended.	Parameter output	OFF	4050
<u> </u>			

-Formatting the GDE alignment using Microsoft Word®

Like for most of sequences analysis, it is necessary to format sequences:

Open the GDE file using a software that is able to read text format (for instance: note pad) Copy then paste the GDE alignment in a Microsoft Word® file

then :

1/ Delete all paragraph jumps:

For instance,

replace:

```
#57337355
------tctgcggcagtgaaagagttcttcggttccagccagctgt
ctcagttcatggaccagaacaacccgctgtctgagatcacgcacaaacgt
cgtatctccgcactcggcccaggcggtctgacccgtgaacgcgcaggctt
#57337357
------aaacgt
cgtatctccgcactcggcccaggcggtctgacccgtgaacgcgcaggctt
```

by:

```
#57337355-----
tctgcggcagtgaaagagttcttcggttccagccagctgtctcagttcatggaccagaacaacccgctg
tctgagatcacgcacaaacgtcgtatctccgcactcggcccaggcggtctgacccgtgaacgcgcaggc
tt#57337357-----
```

aaacgtcgtatctccgcactcggcccaggcggtctgacccgtgaacgcgcaggctt

2/ Replace the « - » by another kind that does not lead to paragraph jump:

For instance,

replace:

#57337355------

a a a cgtcgtatctccgcactcggcccaggcggtctgacccgtgaacgcgcaggctt

by:

```
#57337355.....tctgcggcagtgaaagagttcttcggttccagccagctgtctcagttcat
ggaccagaacaacccgctgtctgagatcacgcacaaacgtcgtatctccgcactcggcccaggcggtct
gacccgtgaacgcgcaggctt#57337357.....aaacgtcgtatctc
cgcactcggcccaggcggtctgacccgtgaacgcgcaggctt
```

3/ Add a paragraph jump before and after each sequence name:

For instance,

replace:

```
#57337355.....tctgcggcagtgaaagagttcttcggttccagccagctgtctcagttcat
ggaccagaacaacccgctgtctgagatcacgcacaaacgtcgtatctccgcactcggcccaggcggtct
gacccgtgaacgcgcaggctt#57337357.....aaacgtcgtatctc
cgcactcggcccaggcggtctgacccgtgaacgcgcaggctt
```

by:

#57337355

```
.....tctgcggcagtgaaagagttcttcggttccagccagctgtctcagttcatggaccagaa
caacccgctgtctgagatcacgcacaaacgtcgtatctccgcactcggcccaggcggtctgacccgtga
acgcgcaggctt
#57337357
.....aaacgtcgtatctccgcactcggcccaggcggtctgacccgtga
acgcgcaggctt
```

The Microsoft word® file is now formatted, and can be copied then pasted into SVARAP

-Pasting the GDE alignment in SVARAP in the Microsoft Excel® File Sheet <u>« Paste the alignment »:</u>

Step 1: When using this program: click on column B then key <Suppr> to delete prior work. Step 2: Paste in a same cell (white space, cell B2) the GDE alignment formatted using Microsoft Word®.



-You can check that the GDE alignment has been well formatted in the Microsoft Excel® Sheet <u>« Sep1000 »:</u>

Check that:

- In column <u>« Identification »:</u> only sequence names are present; check that the number of studied sequences is right.
- In columns <u>« Nucleotides...</u> »: only sequences are present.

If not OK: check the GDE alignment, then paste again.

Identification	Nucleotides 1-1000	Nucleotides 1000-2000	Nucleotides 2000-3000	Nucleotides 3000-4000	<u> </u>
#embl AE0144		gcatggacggcgaagaaattctgtccs	gatacggtcatgccgca	gaacggctacaactacgaagattcgat	
#embl AE0144		gcatggacggcgaagaaattctgtccs	gatacggtcatgccgca	gaacggctacaactacgaagattcgat	
#embl AE0091		gcatggatggcgaagaaatcctgtcga	gacaccgtgatgccgca	gaatggetacaactacgaagaetegat	
#embl BX6404		gcatgacgcccgaatcgatcctcgcgc	gaaaacctgatgccgca	gaacggetacaacttegaagaetegat	
#embl AL6460		gcctgacgccggaacagatcctggcgc	gaaaacctgatgccgca	gaacggctacaacttcgaggattcgat	
#embl AE0169		gctacaccaacgaagagatcctgtccc	gacaacctgatgccgca	gaacggctacaactacgaagactcgat	
#embl AE0048		gctacagcacggaagagatcctcaacg	gaaggeetgatgeegea	gaacggettcaacttegaagaeteeat	
#embl AL1627		gctacaacaatgagcaaatcttggata	gaaaacttgatgccgca	gaacggttacaactatgaagactcgat	-
#embl BX3218		ggtatteteecgcacaaattetegegg	gaaaatcttatgccaca	gaacggacttaatttcgaggattctat	
#embl AE0169		gttatgataccgagcaaatcctcgate	gaaaacctcatgcctca	gaacggatataacttcgaagattccat	
#embl AE0056		actacaccacagagcagatcctcgacc	gataccctgatgcctca	gaatggttacaacttcgaagactccat	
#embl AE0136		atttcaccactgcacagatcttggatc	gatactctgatgccaca	gaacggttacaacttcgaagactccat	
#embl U32733		gctacaccactgaagaaatcttaaact	gatgcgatcacgccaca	gaacggctotoogtttroogestroogt	
#embl AE0062		attacaccaccgaacaaattttagate	gatgcagtgacaccaca	gaatggtt Link to MAIN PAGE t	
#embl AE0140		attataatacagaacaaatattagaca	gacatgettatgeetca	gaatggatataactttgaagattcaat	
#embl AP0025		tccgagctcttggatactcttcage	cctcagatacgttgactccagg	gtatggatacaactttgaggatgcgat	
#embl AE0013		tccgagctcttggatactcttcage	cctcagatacgttgactccagg	gtatggatacaactttgaggatgcgat	
#embl AE0023		ttcgagcattggggtattcgacage	cttccgatacgttaactcctgg	gtatgggtacaacttcgaagatgcgat	
#embl AE0085		g.tatgagtactgaagaaattataaas	gataccgtaatgcctca	gaacggttataattttgaagattccat	
#embl AF0345		g.tatgagtactgaagaaattataaas	gataccgtaatgcctca	gaacggttataattttgaagattcaat	
#embl AL1390		ggtataaaaaacaagatattattaagt	atcttgataaagtaatgcctta	gaatggttataactatgaggatgctat	
#embl AE0011		ggtttgatacgagagaaaaaataatag	aggaagtttttaatctaaagcctca	gaatggatttaattatgaggatgctat	
#embl AE0012		actacgaaacgcgtgagcagatcatcg	aggaaacggaaacgatcaagccgca	gaacggttacaactacgaggatgccat	
#embl AE0085		gtttctcaggtgatgatgaaatctttg	ggacaatgaagtettgacaccaca	ggaaggttacaacttcgaggatgccgt	
#embl AE0064		gtttctcaggtgatgatgaaattgttg	agacaacgatgtgttaacaccaca	ggaaggttataactttgaggatgccgt	
#embl AE0173		getttggtteegateaagaaattattg	agatatgactacaattacaccgca	ggatggttataactacgaggatgcgat	
#embl AP0033		gtttctcaagtgaccaagaaattgttg	agatactgagtctatcacacctca	ggacggttacaactatgaggatgccgt	
#embl AL4455		gttttactgaagaaagcatgtaca	agaaatetetagaateaeteeaaa	aaatggatataactatgaagatgctat	
I I I I I Paste	alignment) Sep1000 / Nuc 1-1000 / Nuc	1001-2000 🖌 Nuc 2001-3000 🖌 Nuc 3001-40	00 🔏 Table sliding wind 25 🔏 Graphe sliding	wind 25 / 4	- I • Ē

-Analysis of variability:

Go back to the **main page (in the Microsoft Excel® File Sheet** <u>« Paste the alignment »)</u> by using any of the boxes called <u>"Link to main page" (hypertext link)</u>. Then select one of the four hypertext links to analysis of variability:



1-Link to analysis of variability by nonsliding windows of 50 nucleotides:



2-Link to site by site analysis of variability:

- The green line represents the absolute number of sequences analysed in the studied set of sequences;
- The blue rods represent variability at each site in the alignment of sequences;
- Nucleotide positions are noted in abcissa:



3-Link to analysis of variability by sliding window of 25 nucleotides:



4-Link to detailed analysis of variability (per 200 nucleotides; click on the Microsoft Excel® sheet name index):

These tables contain the main part of analysis of variability: the level of variability correspond to the proportion of sequences for which, at a given nucleotidic site, the nucleotide differ compared with the nucleotide the most frequently found in the studied set of sequences.

- Positions that are defined correspond to those defined in the ClustalX alignment for the studied set of sequences.
- The number of distinct variations corresponds to the number of different nucleotides observed at a given site.
- Both absolute numbers and proportions of nucleotides harbored at a given site are given.

	D	E	F	G	Н		J	K	L	M	N	0	Р	Q	R	S	Т	U	V	W	Х	Y	Ζ	AA	AB	
105	Positions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
106	Defined positions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
109	Number of sequences	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	
110	Consensus	Α	Т	G	Т	С	Α	С	Α	Α	Α	Α	Α	Α	С	Т	Α	Α	Т	Т	Т	С	Т	Т	Т	
111	Absolute values																									
112	A	2	0	1	1	1	3	0	3	3	2	3	3	1	1	1	2	1	1	0	1	0	0	0	1	-
113	C C	0	1	0	0	2	0	2	0	0	0	0	0	1	2	0	0	0	0	0	0	3	0	0	0	-
114	LINK TO MAIN PAGE T	0	2	0	2	0	0	1	0	0	1	0	0	1	0	2	1	1	2	3	2	0	3	2	2	
115	G	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1
117	Other	33	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	
122	8																									
123	А	6	0	3	3	3	9	0	9	9	6	9	9	3	3	3	6	3	3	0	3	0	0	0	3	ī.
124	с	0	3	0	0	6	0	6	0	0	0	0	0	3	6	0	0	0	0	0	0	9	0	0	0	1
125	т	0	6	0	6	0	0	3	0	0	3	0	0	3	0	6	3	3	6	9	6	0	9	6	6	1
126	G	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	3	0	Ī
128	Other	94	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	
136	Nb of distinct variations	1	2	2	2	2	1	2	1	1	2	1	1	3	2	2	2	3	2	1	2	1	1	2	2	
138	Variability	6	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	
139	Positions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
140	Defined positions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
141																										
142																										
143																										
144																										
145																										
146																										
147																										
148																										
14 4	▶ N / Data window 50 / Data 50_N	IOT IN	USE /	Fig 1	-4000 v	vindow	50 /	Consen	sus 🗸 i	Consen	sus (2)	1×1	(2) <i>k</i>	v X v G	2)), Ya	r 1-20	0 / V.	ar 201-4	400 X							Ē

All figures or tables are printable.

An example of the use of SVARAP: Investigation of the usefulness of *rpoB* sequencing for differentiation and identification of bacteria belonging to the genera *Afipia* and *Bosea* which are amoeba-resisting bacteria recently reported to colonize hospital water supplies (<u>Appl</u> Environ Microbiol 2003, 69(11):6740-9):

SVARAP: Sequence VARiability Analysis Program

An example of application: rpoB gene sequencing for identification of Corynebacterium species

```
Clustal X alignment
```

Khamis, A., Raoult, D. & La Scola, B. (2004). J Clin Microbiol 42:3925-31.



Corynebacterium species: variability in rpoB

« One shot » vizualisation of nucleotide variability (nonsliding window of 50 nucleotides)



« One shot » vizualisation of nucleotide site-by-site variability (nuc. 2001-4000 of the alignment)



Corynebacterium species: variability in rpoB



« One shot » vizualisation of nucleotide site-by-site variability (nuc. 2001- 4000 of the alignment)

Vizualisation of nucleotide site-by-site variability: primers hibridization regions





	Forward primer														_									
Defined positions	2736	2737	2738	2739	2740	2741	2742	2743	2744	2745	2746	2747	2748	2749	2750	2751	2752	2753	2754	2755	2756	2757	2758	2759
Number of sequences	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58
Consensus	Т	С	G	т	Α	т	G	Α	А	С	Α	т	С	G	G	С	С	Α	G	G	т	G	С	Т
Absolute values																								
А	0	0	0	2	58	0	0	58	58	0	58	0	0	0	0	0	0	58	1	0	0	0	0	0
Link to MAIN PAGE	8	58	0	1	0	0	0	0	0	58	0	0	43	0	0	43	58	0	0	0	0	21	57	0
Т	50	0	0	55	0	58	0	0	0	0	0	58	15	0	0	- 14	0	0	0	0	58	15	1	58
G	0	0	58	0	0	0	58	0	0	0	0	0	0	58	58	1	0	0	-57	58	0	22	0	0
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	3	100	0	0	10 0	100	0	100	0	0	0	0	0	0	10 0	2	0	0	0	0	0
c	- 14	100	0	2	0	0	0	0	0	100	0	0	74	0	0	74	100	0	0	0	0	36	98	0
т	86	0	0	95	0	100	0	0	0	0	0	100	26	0	0	24	0	0	0	0	100	26	2	100
a	0	0	100	0	0	0	100	0	0	0	0	0	0	100	100	2	0	0	- 98	100	0	38	0	0
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nili af distinct vecisticans	1	0	0	2	0	0	0	0	0	0	0	0	1	0	0	2	0	0	1	0	0	2	1	0
Variability	14	0	0	5	0	0	0	0	0	0	0	0	26	0	0	26	0	0	2	0	0	62	2	0
	•																							

Corynebacterium species: variability in rpoB ClustalX (v1.8) COTTOT CACCATE •SVARAP: analysis of variability per site and per sliding window of 25 nucleotides **Reverse primer** 70 60 50 40 30 20 10 Number of sequences 50 40 30 20 10 Variability (%) llmII Ō Proportion of sequences harbouring a nucleotide at a given site Variability (%)

	Reverse primer																							
Defined positions	3167	3168	3169	3170	3171	3172	3173	3174	3175	3176	3177	3178	3179	3180	3181	3182	3183	3184	3185	3186	3187	3188	3189	3190
Number of sequences	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58
Consensus	С	С	С	Α	G	С	G	С	Т	т	С	G	G	С	G	Α	G	А	т	G	G	А	G	G
Absolute values																								
A	0	6	0	58	0	0	0	0	0	0	0	0	0	1	0	58	9	58	0	0	0	58	3	0
Link to MAIN PAGE	56	43	57	0	0	58	0	51	0	0	56	0	0	31	0	0	0	0	0	0	0	0	0	0
Т	2	9	0	0	0	0	0	7	58	58	2	0	0	26	0	0	0	0	58	0	0	0	0	2
G	0	0	1	0	58	0	58	0	0	0	0	58	-58	0	58	0	49	0	0	58	58	0	6 5	56
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8																								
A	0	10	0	100	0	0	0	0	0	0	0	0	0	2	0	100	16	100	0	0	0	100	5	0
c	97	74	98	0	0	100	0	88	0	0	-97	0	0	53	0	0	0	0	0	0	0	0	0	0
т	3	16	0	0	0	0	0	12	100	100	3	0	0	45	0	0	0	0	100	0	0	0	0	3
G	0	0	2	Q	100	0	100	0	0	0	0	100	100	0	100	0	84	0	0	100	100	0	95	97
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alt of distinct variations	1	2	1	0	0	0	0	1	0	0	1	0	0	2	0	0	1	0	0	0	0	0	1	1
Variability	3	26	2	Ō	Ō	Ō	0	12	Ō	0	3	Ō	Ō	47	Ō	Ō	16	Ō	Ō	Ō	Ō	0	5	3

SVARAP download :

- Philippe Colson, Unité des Rickettsies, CNRS UMR 6020, IFR 48, Faculté de Médecine, Université de la Méditerranée, Marseille, France.
- http://ifr.free.fr/recherche/jeu_cadre/jeu_rickettsie.html
- · Hypertext link « Téléchargement »

aSVARAP user manual

The study of variability can also concern amino acid sequences (amino acids 1 to 1000). The principle and use are the same as for SVARAP.

References for published studies using SVARAP or aSVARAP

- 1. Adékambi T, Colson P, Drancourt M. *Rpo*B-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. J Clin Microbiol 2003, 41(12):5699-5708.
- Colson, P., M. Henry, C. Tourres, D. Lozachmeur, H. Gallais, J. A. Gastaut, J. Moreau, and C. Tamalet. 2004. Polymorphism and drug-selected mutations in the protease gene of human immunodeficiency virus type 2 from patients living in Southern France. <u>J Clin</u> <u>Microbiol 2004</u>, 42(2):570-577.
- Colson P, Henry M, Tivoli N, Gallais H, Gastaut JA, Moreau J, Tamalet C. Polymorphism and drug-selected mutations in the reverse transcriptase gene of HIV-2 from patients living in southeastern France. <u>J Med Virol 2005</u>, 75: 381-390.

- Khamis A, Colson P, Raoult D, La Scola B. Usefulness of rpoB gene sequencing for identification of Afipia and Bosea species including a strategy for the choice of discriminative partial sequences. <u>Appl Environ Microbiol 2003</u>, 69(11):6740-9.
- Khamis A, Raoult D, La Scola B. rpoB gene sequencing for identification of Corynebacterium species. J Clin Microbiol. 2005 Apr;43(4):1934-6.

Other references

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876-4882.

Contact

Philippe Colson

Fédération Hospitalière de Microbiologie Clinique et d'Hygiène, CHRU Timone, Laboratoire de Virologie, 264 rue Saint-Pierre 13385, Marseille Cedex 05.

Phone number: +33 (0)4 91 38 55 19 ; Fax number: +33 (0)4 91 38 55 18; E-mail address: Philippe.Colson@ap-hm.fr

Unité des Rickettsies, CNRS UMR 6020 IFR48, Faculté de Médecine, Université de la Méditerranée, 27 boulevard Jean Moulin, 13 385 Marseille Cedex 05. Phone number: +33 (0)4 91 38 55 17.

Downloads

SVARAP and aSVARAP: <u>http://ifr48.timone.univ-mrs.fr/portail2/index.php?option=com_content&task=view&id=50 -</u> <u>telechargements</u> An example of use for SVARAP: <u>http://ifr48.timone.univ-mrs.fr/files/SVARAP/ExampleUseSVARAP.pdf</u>