Appendix 1.
Frequencies of putative strong promoters observed over 32 prokaryotic genomes

n	a	s (Mbp)	at	g	CI		CII	
					sp_{CI}	$p1_{CI}$ (%)	sp_{CII}	$p1_{CII}$ (%)
Aquifex aeolicus vf5	Others_AA	1.55	57.62	**1522**	**214**	14.1	**333**	21.9
Bacillus subtilis 168	Firm_BS	4.21	59.27	**3979**	**796**	20.0	**1225**	30.8
Borrelia burgdorferi b31	Spiro_BB	0.91	73.07	**850**	**43**	5.1	**49**	5.8
Brucella melitensis 16m chr1	Proteo_BM	3.28	45.56	**2059**	**25**	1.2	**93**	4.5
Chlamydophila pneumoniae ar 39	Chla_CPn	1.22	62.55	**1069**	**23**	2.2	**30**	2.8
Clostridium perfringens str13	Firm_CPe	3.03	74.93	**2532**	**797**	31.5	**835**	33.0
Deinococcus radiodurans r1 chr1	Others_DR	3.05	34.01	**2521**	**22**	0.9	**204**	8.1
Escherichia coli k12	Proteo_EC	4.63	52.34	**4173**	**96**	2.3	**254**	6.1
Haemophilus influenza rd kw20	Proteo_HI	1.83	64.22	**1673**	**31**	1.9	**37**	2.2
Helicobacter pylori j99	Proteo_HP	1.64	62.95	**1478**	**31**	2.1	**34**	2.3
Listeria innocua	Firm_LI	3.01	64.25	**2962**	**713**	24.1	**946**	31.9
Listeria monocytogenes strain EGD	Firm_LM	2.94	63.97	**2837**	**707**	24.9	**926**	32.6
Mycobacterium leprae tn	Atb_ML	3.26	49.55	**2670**	**31**	1.2	**122**	4.6
Mycobacterium tuberculosis h37rv	Atb_MT	4.41	35.47	**3909**	**32**	0.8	**290**	7.4
Mycoplasma genitalium G37	Molli_MGe	0.58	69.54	**441**	**4**	0.9	**4**	0.9
Mycoplasma pneumoniae M129	Molli_MPn	0.81	61.74	**644**	**18**	2.8	**26**	4.0
Neisseria meningitidis mc58	Proteo_NM	2.27	51.90	**1954**	**50**	2.6	**92**	4.7
Oceanobacillus iheyensis hte831	Firm_OI	3.63	66.41	**3398**	**882**	26.0	**1049**	30.9
Pseudomonas aeruginosa pa01	Proteo_PAe	6.26	35.28	**5565**	**45**	0.8	**353**	6.3
Rickettsia prowazekii madrid e	Proteo_RPM	1.11	73.16	**796**	**6**	0.8	**6**	0.8
Salmonella typhimurium lt2	Proteo_ST	4.85	51.22	**4334**	**117**	2.7	**342**	7.9
Shewanella oneidensis mr1	Proteo_SO	4.96	56.69	**4501**	**118**	2.6	**167**	3.7
Sinorhizobium meliloti 1021	Proteo_SM	3.65	39.50	**3272**	**55**	1.7	**385**	11.8
Staphylococcus aureus mw2	Firm_SA	2.82	69.77	**2610**	**560**	21.5	**622**	23.8
Streptococcus pneumoniae r6	Firm_SPn	2.03	62.39	**1861**	**213**	11.5	**285**	15.3
Streptomyces coelicolor a3 (2)	Atb_SC	8.66	29.15	**4665**	**22**	0.5	**347**	7.4
Thermoanaerobacter tengcongensis	Firm_TT	2.68	63.83	**2588**	**581**	22.5	**715**	27.6
Thermotoga maritima	Others_TM	1.86	54.59	**1790**	**305**	17.0	**707**	39.5
Treponema pallidum nichols	Spiro_TPN	1.13	47.00	**980**	**41**	4.2	**125**	12.8
Vibrio cholerae n16961 chr1	Proteo_VC	4.03	54.80	**2618**	**30**	1.2	**72**	2.8
Xanthomonas campestris atcc 33913	Proteo_XC	5.07	35.55	**4120**	**6**	0.2	**97**	2.4
Yersinia pestis	Proteo_YP	4.6	55.36	**4090**	**52**	1.3	**136**	3.3

Table 1.1 Frequencies of genes harbouring a potentially strong $\sigma 70$ promoter, under two constraint sets, in 32 prokaryotic genomes. n: micro-organism name; a: abbreviation for micro-organism name (Atb: *Actinobacteria*, Chla: *Chlamydia*, Firm: *Firmicutes* (among which Molli.: *Mollicutes*), "Others" group, Proteo: *Proteobacteria*, Spiro: *Spirochaetales*); s: genome size; at: average percentage of nucleotides T and A over the 350 bp-long region upstream of start codon, computed over the total number g of genes encoding proteins in the genome; sp_{CI}: number of genes harbouring a putative **S**trong **P**romoter identified under constraint set CI, sp_{CII}: *idem*, under the more relaxed constraint set CII (see text, Subsection "Genome analysis upon request" for the definition of CI and CII constraints); $p1_{CI} = 100 \times sp_{CI}/g$, $p1_{CII} = 100 \times sp_{CII}/g$.

n	a	at	g	CI		CII		$p1_{CI} \times p2_{CI}$ (%)	$p1_{CII} \times p2_{CII}$ (%)
				$upsp_{CI}$	$p2_{CI}$ (%)	$upsp_{CII}$	$p2_{CII}$ (%)		
Aquifex aeolicus vf5	Others_AA	57.62	**1522**	**19**	8.9	**73**	21.9	1.3	4.8
Bacillus subtilis 168	Firm_BS	59.27	**3979**	**115**	14.5	**416**	34.0	2.9	10.5
Borrelia burgdorferi b31	Spiro_BB	73.07	**850**	**25**	58.1	**40**	81.6	2.9	4.7
Brucella melitensis 16m	Proteo_BM	45.56	**2059**	**0**	0	**4**	4.3	0	0.2
Chlamydophila pneumoniae ar 39	Chla_CPn	62.55	**1069**	**6**	26.1	**14**	46.7	0.6	1.3
Clostridium perfringens str13	Firm_CPe	74.93	**2532**	**511**	64.1	**730**	87.4	20.2	28.8
Deinococcus radiodurans r1	Others_DR	34.01	**2521**	**0**	0	**1**	0.5	0	0.0
Escherichia coli k12	Proteo_EC	52.34	**4173**	**3**	3.1	**22**	8.7	0.1	0.5
Haemophilus influenza rd kw20	Proteo_HI	64.22	**1673**	**8**	25.8	**20**	54.1	0.5	1.2
Helicobacter pylori j99	Proteo_HP	62.95	**1478**	**7**	22.6	**17**	50.0	0.5	1.2
Listeria innocua	Firm_LI	64.25	**2962**	**145**	20.3	**501**	53.0	4.9	16.9
Listeria monocytogenes strain EGD	Firm_LM	63.97	**2837**	**147**	20.8	**488**	52.7	5.2	17.2
Mycobacterium leprae tn	Atb_ML	49.55	**2670**	**0**	0	**1**	0.8	0	0.0
Mycobacterium tuberculosis h37rv	Atb_MT	35.47	**3909**	**0**	0	**2**	0.7	0	0.1
Mycoplasma genitalium G37	Molli_MGe	69.54	**441**	**1**	25.0	**1**	25.0	0.2	0.2
Mycoplasma pneumoniae M129	Molli_MPn	61.74	**644**	**2**	11,1	**10**	38.5	0.3	1.6
Neisseria meningitidis mc58	Proteo_NM	51.90	**1954**	**3**	6,0	**13**	14.1	0.2	0.7
Oceanobacillus iheyensis hte831	Firm_OI	66.41	**3398**	**217**	24.6	**576**	54.9	6.4	17.0
Pseudomonas aeruginosa pa01	Proteo_PAe	35.28	**5565**	**0**	0	**6**	1.7	0	0.1
Rickettsia prowazekii madrid e	Proteo_RPM	73.16	**796**	**2**	33.3	**6**	100	0.3	0.8
Salmonella typhimurium lt2	Proteo_ST	51.22	**4334**	**10**	8.6	**40**	11.7	0.2	0.9
Shewanella oneidensis mr1	Proteo_SO	56.69	**4501**	**10**	8.5	**38**	22.8	0.2	0.8
Sinorhizobium meliloti 1021	Proteo_SM	39.50	**3272**	**0**	0	**6**	1.6	0	0.2
Staphylococcus aureus mw2	Firm_SA	69.77	**2610**	**225**	40.2	**431**	69.3	8.6	16.5
Streptococcus pneumoniae r6	Firm_SPn	62.39	**1861**	**59**	27.7	**120**	42.1	3.2	6.5
Streptomyces coelicolor a3 (2)	Atb_SC	29.15	**4665**	**0**	0	**0**	0	0	0
Thermoanaerobacter tengcongensis	Firm_TT	63.83	**2588**	**150**	25.8	**407**	56.9	5.8	15.7
Thermotoga maritima	Others_TM	54.59	**1790**	**31**	10.2	**120**	17.0	1.7	6.7
Treponema pallidum nichols	Spiro_TPN	47.00	**980**	**1**	2.4	**8**	6.4	0.1	0.8
Vibrio cholerae n16961	Proteo_VC	54.80	**2618**	**2**	6.7	**5**	6.9	0.1	0.2
Xanthomonas campestris atcc 33913	Proteo_XC	35.55	**4120**	**0**	0	**0**	0	0	0
Yersinia pestis	Proteo_YP	55.36	**4090**	**3**	5.8	**20**	14.7	0.1	0.5

Table 1.2 Frequencies of genes with a putative $\sigma 70$ promoter harbouring an UP element, under two constraint sets, in 32 prokaryotic genomes. n micro-organism name; a: abbreviation for micro-organism name (see Table 1.1.); at: average percentage of nucleotides T and A over the 350 bp-long region upstream of start codon, computed over the total number g of genes in the genome; $upsp_{CI}$: number of genes with an **UP** element harboured in the **S**trong **P**romoter region, identified under constraint set CI; $upsp_{CII}$: *idem*, under the more relaxed constraint set CII (see text, Section "Genome analysis upon request" for the definition of CI and CII constraints); $p2_{CI} = 100 \times upsp_{CI}/sp_{CI}$, $p2_{CII} = 100 \times upsp_{CII}/sp_{CII}$, with sp_{CI} (resp. sp_{CII}) the number of genes harbouring a **S**trong **P**romoter under constraints CI (resp. CII). The percentages indicated in the two rightest columns are respectively $p1_{CI} \times p2_{CI} = upsp_{CI}/g$ and $p1_{CII} \times p2_{CII} = upsp_{CII}/g$, that is the ratio of the number of genes identified with an UP element in their putative strong promoter, to the number g of genes encoding proteins in the genome considered.