

Table S4 Motif distribution in invertase family

Invertase Family	Motif number	Length (aa)	Consensus sequence	E value
Acid invertase	1	34	RRILWGWANESDSEADDVAKGWAGIQAIPTIWL	9.3e-537
	2	23	NYQVQNIAFPKNPSDPLLREWVK	4.1e-437
	3	41	WGNIVWGHSVSTD LINWIALEPAIEPTIPFDINGCWSGSAT	7.7e-384
	4	36	TISLRTLIDHSIVESFGQGGRTCIMSRVYPEHAITD	8.3e-433
	5	29	HGGVGPFGWLWVMASGNMREHTAVFFRVFK	3.2e-410
	6	21	DPDGKQLLQWPVEEIIETLRGK	7.3e-307
	7	15	YYKGMYPHFYQYNPH	5.8e-316
	8	29	LMCTDPTKSSTKPNLYKPTYGGFVDVDIE	2.0e-287
	9	21	PIPPPQGINANQFRDPTTAW	1.4e-293
	10	21	WLRDYGNFYASKTFFDPAKR	6.3e-267
	11	29	AHMYVFNNGTDDVKVSKLEAWEMKKAQMN	6.8e-255
	12	21	SPHYRTAYHFQPPKNWINDPN	2.9e-205
	13	15	SNATGMWECPDFFPV	1.4e-236
	14	21	TRCDYYMIGTYDDKTDTFVPD	3.9e-180
	15	29	SAEVDGYGAAVYRSRDFKHWERVRHPLH	4.4e-112
Neutral/alkaline			EMPLKICYPAMENHEWQIVTGCDPKNTRWSYHNGGSWPVLLWQLTAACI	
invertase	1	50	K	2.5e-736
	2	50	SYHMRSYYWLDLDFQQLNDIYRYKTEEYSHTAVNKFNVYPEQIPDWLDFDFMP	2.4e-699
	3	50	RLILNLCLSEGFDTFPTLLCTDGCCMIDRRMGIYGYPIEQALFYMALRC	4.9e-674
	4	50	NYDQVFVRDFVPSALAFMNGETEIVKNFLLHTLQLQSWEKMVDCFKLGQ ARRAIEVAEQRLAKDKWPEYYDGKLGRYVGKQARKFQTWSIAGYLVAK	3.3e-639
	5	50	MM	2.7e-607
	6	50	ETLVADFGESAIGRVAPVDSGFWWIILLRAYTKWTGDNLSLAETPDCQKGM	6.9e-607

7	50	GGYFIGNVQPAHMDFRWFCLGNCWAIVSSLATPEQAEAIMDLIEEKWDDL	7.7e-591
8	50	HTPASTHFNSFEPHPIVNEAWEALRRSVVYFCGQPVGTVAAVDHSCEEVL	2.6e-346
9	21	MLKPQNEGKDFIEHINTRLHA	4.0e-122
10	15	GAMPASFKVHHNPKD	4.6e-122
11	21	LEDPSHLM MISCEEDKHVKPP	1.3e-104
12	21	QRINVERQRSFDDRSWSEL SH	7.0e-88
13	50	DENHDSL PKPEKRTRMHHIERHRSCVVTLS DMELNGLQPRLLQTIEKSP	5.2e-61
14	50	KTAAGTATRLFPRLPHPHPPYGRRLPFLVSAASQS QASQTGPATPVPVP	4.3e-58
15	21	MRRVSSHVSLADPDDFNLTRL	9.4e-42

---