

Лабораторная работа 4

Работа выполнялась на ПК 16Gb RAM, Apple M2 Pro, с предустановленной MacOS Sonoma 14.4.1.

Цель работы: знакомство с распознаванием пептидов по из масспектральному отпечатку при помощи программы Cruх.

1. Бинарные файлы Cruх (версия OS X) были загружены с официального сайта <http://cruх.ms>, версия 4.2.
2. Необходимые для распознавания данные (непосредственно файл результатов масс-спектрометрии и небольшая база данных белков) также были взяты с официального сайта и помещены в директорию data.
3. Индексирование входных данных было произведено командой

```
cruх tide-index small-yeast.fasta yeast-index
```

4. Ожидалось, что в результате выполнения будут получены директории cruх-output и yeast-index. Однако содержимое директории cruх-output не совпало с указанным на сайте:
 - 1) tide-index.decoy.fasta – множество найденных белков из входного множества
 - 2) tide-search.params.txt – запись параметров, использованных при поиске
 - 3) tide-search.log.txt – файл логированияПри выполнении команды tide-index не был создан файл tide-index.decoy.fasta.
5. Поиск и ранжирование пептидов, чьи ожидаемые массы близки к полученным в ходе масс-спектрометрии, были произведены командой

```
cruх tide-search demo.ms2 yeast-index
```

6. В результате были получены файлы

```

maria@Marias-MacBook-Pro lab4 % ls -l crux-output
total 2816
-rw-r--r--  1 maria  staff      835 Apr 25 08:43 tide-index.log.txt
-rw-r--r--  1 maria  staff  89064 Apr 25 08:43 tide-index.params.txt
-rw-r--r--  1 maria  staff 611328 Apr 25 08:49 tide-search.decoy.txt
-rw-r--r--  1 maria  staff   1397 Apr 25 08:49 tide-search.log.txt
-rw-r--r--  1 maria  staff  89064 Apr 25 08:49 tide-search.params.txt
-rw-r--r--  1 maria  staff 540547 Apr 25 08:49 tide-search.target.txt

```

7. Отсортировав файл tide-search.target.txt по убыванию XCorr, получили

1	file	scan	charge	spectrum precursor m/z	spectrum neutral mass	peptide mass	delta_cn	delta_ion	xcorr score	b/y ions total	b/y ions fraction	xcorr rank	distinct matches/spectrum	sequence	modifications	unmodified sequence	protein id	flanking aa	target
2	demo.ms2	60135	3	1057.8792	3170.6156	3170.6110	0	0	6.69412737	42	91	0.46153846	1	1	ALSRPNVEVALNDPFI	ALSRPNVEVALNDPFI	YGR182C19	RY	target
3	demo.ms2	60355	4	838.1677	3348.6418	3348.6414	0	0	6.36378876	50	95	0.52631579	1	1	HEASEVASFLNGNIEHVP	HEASEVASFLNGNIEH	YL9249W27	RG	target
4	demo.ms2	57701	3	1190.5835	3568.7287	3568.7209	0	0	6.23474770	41	108	0.37962963	1	1	GVLYTEDAVSSDFLGDH	GVLYTEDAVSSDFLG	YGR182C278	KF	target
5	demo.ms2	44517	3	726.3629	2215.0698	2215.0608	0	0	6.11194608	42	75	0.50000000	1	1	HELSSADYVNDQGTGHR	HELSSADYVNDQGTG	YCR029W153	RA	target
6	demo.ms2	75478	3	975.1579	2922.4519	2922.4466	0	0	5.96729623	35	98	0.35714286	1	1	NMTGTSGQ/11_S_57.0215	NMTGTSGQACALJAG	YBR1189U101	KD	target
7	demo.ms2	57996	3	1087.8482	3260.5167	3260.5083	0	0	5.89492096	49	98	0.49494949	1	1	LGLGGGDMPSSEADPVE	LGLGGGDMPSSELA	YGL080C584	RY	target
8	demo.ms2	71336	3	852.4749	2554.4029	2554.3927	0	0	5.86348290	36	90	0.40000000	1	1	TQDILLDAPLSLGIETAGI	TQDILLDAPLSLGIET	YL924C289	KL	target
9	demo.ms2	64346	3	859.1065	2574.2977	2574.2999	0	0	5.75522178	36	84	0.42857143	1	1	VINDAFIEEGLMTTHSLTAI	VINDAFIEEGLMTTH	YGR182C161	KT	target
10	demo.ms2	58198	3	1190.5835	3568.7287	3568.7209	0	0	5.65417739	38	108	0.35185185	1	1	GVLYTEDAVSSDFLGDH	GVLYTEDAVSSDFLG	YGR182C278	KF	target
11	demo.ms2	58336	3	935.1379	2802.3918	2802.3868	0	0	5.65212335	37	86	0.43023256	1	1	STLUNLVLTGE 24_S_57.0215	STLUNLVLTGELTSGEV	YL9249W108	KI	target
12	demo.ms2	57206	3	1190.5833	3568.7279	3568.7209	0	0	5.61496434	38	108	0.35185185	1	1	GVLYTEDAVSSDFLGDH	GVLYTEDAVSSDFLG	YGR182C278	KF	target
13	demo.ms2	34538	3	673.6952	2018.0637	2018.0618	0	0	5.56736838	40	80	0.50000000	1	1	LPVVGHEGAGVWGMGEN	LPVVGHEGAGVWGM	YMR303C81	KG	target
14	demo.ms2	55157	3	982.4869	2944.4388	2944.4310	0	0	5.55117069	39	89	0.43620225	1	1	LSLHEPYTV 22_S_57.0215	LSLHEPYTVGDSVPOD	YGL080C388	KK	target
15	demo.ms2	48115	4	554.7750	2215.0710	2215.0658	0	0	5.47434758	34	75	0.45333333	1	1	HELSSADYVNDQGTGHR	HELSSADYVNDQGTG	YCR029W153	RA	target
16	demo.ms2	60341	3	869.5145	2875.5217	2875.5178	0	0	5.45826697	30	87	0.34477591	1	1	THLMEALDLELGGTAK	THLMEALDLELGGT	YJ996C238	KG	target
17	demo.ms2	34982	3	673.6962	2018.0669	2018.0618	0	0	5.44820059	39	80	0.48750000	1	1	LPVVGHEGAGVWGMGEN	LPVVGHEGAGVWGM	YMR303C81	KG	target
18	demo.ms2	60170	3	1190.5815	3568.7228	3568.7209	0	0	5.41170391	31	108	0.28703704	1	1	GVLYTEDAVSSDFLGDH	GVLYTEDAVSSDFLG	YGR182C278	KF	target
19	demo.ms2	20706	3	520.9549	1559.8429	1559.8419	0	0	5.40655332	36	56	0.46285714	1	1	SHINWVGHVDSGK	SHINWVGHVDSGK	YBR1189W6	KS	target
20	demo.ms2	59197	3	1190.5833	3568.7279	3568.7209	0	0	5.32893316	31	108	0.28703704	1	1	GVLYTEDAVSSDFLGDH	GVLYTEDAVSSDFLG	YGR182C278	KF	target
21	demo.ms2	22958	3	520.9552	1559.8438	1559.8419	0	0	5.30714966	32	56	0.57142857	1	1	SHINWVGHVDSGK	SHINWVGHVDSGK	YBR1189W6	KS	target
22	demo.ms2	60831	4	838.1685	3348.6450	3348.6414	0	0	5.27053110	36	95	0.37894737	1	1	HEASEVASFLNGNIEHVP	HEASEVASFLNGNIEH	YL9249W27	RG	target
23	demo.ms2	28906	3	731.0362	2190.0868	2190.0817	0.957126	0.957126	5.26683248	32	76	0.42105263	1	2	HLVHEVTSPOAFEGLENAGR	HLVHEVTSPOAFEGLEN	YGL090C37	RK	target
24	demo.ms2	53937	3	746.3829	2236.1268	2236.1244	0	0	5.23052371	43	82	0.52439024	1	1	THLMSPMAMAAAGHGFVC	THLMSPMAMAAAGH	YGL090C458	RE	target
25	demo.ms2	62316	3	974.1095	2919.3067	2919.3061	0	0	5.21462025	32	83	0.38554217	1	1	GPTTFWDGSEFFDSHGLT	GPTTFWDGSEFFDSH	YJ996C147	RD	target
26	demo.ms2	59904	3	859.7819	2576.2638	2576.2605	0	0	5.15799092	31	82	0.33999992	1	1	SINPDEAVYGAQVQALLTG	SINPDEAVYGAQVQ	YL924C239	RT	target
27	demo.ms2	21124	3	520.9549	1559.8429	1559.8419	0	0	5.12927237	33	56	0.58689771	1	1	SHINWVGHVDSGK	SHINWVGHVDSGK	YBR1189W6	KS	target
28	demo.ms2	59393	3	923.8232	2768.4477	2768.4418	0	0	5.09869067	23	89	0.27078652	1	1	LLSQLEGIIP 22_S_57.0215	LLSQLEGIIPLESSHAF	YGL026C398	KT	target
29	demo.ms2	23427	3	520.9549	1559.8429	1559.8419	0	0	5.09859544	32	56	0.57142857	1	1	SHINWVGHVDSGK	SHINWVGHVDSGK	YBR1189W6	KS	target
30	demo.ms2	25798	3	643.3115	1928.9127	1928.9105	0.94277	0.94277	5.07885302	29	64	0.45312500	1	2	VDCLATVQ 3_S_57.0215	VDCLATVQHPTESR	YGL090C81	RK	target
31	demo.ms2	64798	3	859.1075	2574.3006	2574.2999	0	0	5.04052003	31	84	0.36904762	1	1	VINDAFIEEGLMTTHSLTAI	VINDAFIEEGLMTTH	YGR182C161	KT	target
32	demo.ms2	58701	3	956.1359	2865.3859	2865.3814	0	0	5.04265503	33	91	0.36263738	1	1	VLNANTNAPFVSEVMNDVC	VLNANTNAPFVSEVM	YPL106C277	KE	target
33	demo.ms2	22485	3	520.9555	1559.8447	1559.8419	0	0	5.01954306	31	56	0.55357143	1	1	SHINWVGHVDSGK	SHINWVGHVDSGK	YBR1189W6	KS	target
34	demo.ms2	38583	3	816.7822	2447.2587	2447.2543	0	0	5.01612018	31	82	0.37804878	1	1	QSIIEAFGKPLSTLNGDEAL	QSIIEAFGKPLSTLNG	VPL106C352	KG	target
35	demo.ms2	41588	3	816.7822	2429.3249	2429.3239	0	0	5.00647819	34	88	0.38636364	1	1	VETGKPKGMVITFAPAGVTT	VETGKPKGMVITFAPA	YBR1189U101	RS	target
36	demo.ms2	75497	2	1482.2341	2922.4537	2922.4466	0	0	4.99944533	29	42	0.69047619	1	1	NMTGTSGQ/11_S_57.0215	NMTGTSGQACALJAG	YBR1189U101	KD	target
37	demo.ms2	31398	3	585.6475	1753.9207	1753.9210	0	0	4.99885585	34	68	0.50000000	1	1	ISHVSTGGAGLLELEG	ISHVSTGGAGLLELEG	YCR029W387	KE	target

8. Пост-процессинг (поиск лучших результатов для каждого масс-спектра и фильтрация от шумов) был произведён командой

```
crux percolator --test-fdr 0.1 crux-output/tide-search.target
```

9. Отсортировав по убыванию score таблицу совпадений пептида с масс-спектрами, получили

percolator.target.psms						
PSMId	filename	score	q-value	posterior_error_prob	peptide	proteinIds
target_0_60355_4_1	demo.ms2	4.43187	0.00220751	1.41026E-08	R.HEIASEVASFLNGNIEHDVPEHFFGELAK.G	YLR249W
target_0_46517_3_1	demo.ms2	3.9342	0.00220751	8.67569E-08	R.HELSSLADVYINDAFGTAHRA.A	YCR012W
target_0_46115_4_1	demo.ms2	3.8055	0.00220751	1.38787E-07	R.HELSSLADVYINDAFGTAHRA.A	YCR012W
target_0_60135_3_1	demo.ms2	3.70849	0.00220751	1.97767E-07	R.IALSRPNVEVALNDPFITNDYAAVMFK.Y	YGR192C
target_0_71336_3_1	demo.ms2	3.65244	0.00220751	2.42667E-07	K.TQDLLLLDVAPLSLGIETAGVMTK.L	YLL024C
target_0_75478_3_1	demo.ms2	3.64947	0.00220751	2.45319E-07	K.NMITGTQADCAILIIAGGVGEFEAGISK.D	YBR118W
target_0_67696_3_1	demo.ms2	3.55492	0.00220751	3.46444E-07	R.LGLGGGDMPGSELADFVENADGFAEVFPQHK.Y	YGL008C
target_0_64346_3_1	demo.ms2	3.53121	0.00220751	3.77758E-07	K.VINDAFGIEEGLMTTVHSLTATQK.T	YGR192C
target_0_34982_3_1	demo.ms2	3.31474	0.00220751	8.32554E-07	K.LPLVGGHEGAGVWVGMEVVK.G	YMR303C
target_0_58198_3_1	demo.ms2	3.28166	0.00220751	9.39417E-07	K.GVLGYTEDAVVSSDFLGDSSHSSIFDASAGIQLSPK.F	YGR192C
target_0_57206_3_1	demo.ms2	3.24068	0.00220751	1.09101E-06	K.GVLGYTEDAVVSSDFLGDSSHSSIFDASAGIQLSPK.F	YGR192C
target_0_26779_4_1	demo.ms2	3.2361	0.00220751	1.10939E-06	K.SVEMHHEQLQEGVPGDNVGFNVK.N	YBR118W
target_0_60041_3_1	demo.ms2	3.23338	0.00220751	1.12048E-06	K.TIHLNEALDEELQDQDLNAILAK.G	YJR069C
target_0_25798_3_1	demo.ms2	3.23183	0.00220751	1.12683E-06	R.VDCTLATVDHNIPTESR.K	YGL009C
target_0_37063_2_1	demo.ms2	3.12627	0.00220751	1.65656E-06	R.TGDLATGATTSFTEAVIK.R	YOR136W
target_0_40061_5_1	demo.ms2	3.11042	0.00220751	1.75526E-06	K.GCASFNATVVLNHPGQISAGYSPVLDCHTAHIACR.F	YBR118W
target_0_26339_4_1	demo.ms2	3.08467	0.00220751	1.92824E-06	K.SVEMHHEQLQEGVPGDNVGFNVK.N	YBR118W
target_0_53937_3_1	demo.ms2	3.08354	0.00220751	1.93623E-06	R.THLMSPAMAAAAGIAGHFVDIR.E	YGL009C
target_0_23427_3_1	demo.ms2	3.06681	0.00220751	2.05817E-06	K.SHINVVVIGHVDSGK.S	YBR118W
target_0_60170_3_1	demo.ms2	3.02918	0.00220751	2.36123E-06	K.GVLGYTEDAVVSSDFLGDSSHSSIFDASAGIQLSPK.F	YGR192C
target_0_22485_3_1	demo.ms2	3.01279	0.00220751	2.50686E-06	K.SHINVVVIGHVDSGK.S	YBR118W
target_0_59904_3_1	demo.ms2	2.95474	0.00220751	3.09856E-06	R.SINPDEAVAYGAAVQAAILTGEDESSK.T	YLL024C
target_0_20266_3_1	demo.ms2	2.93308	0.00220751	3.35346E-06	K.SHINVVVIGHVDSGK.S	YBR118W
target_0_58336_3_1	demo.ms2	2.92441	0.00220751	3.4613E-06	K.STLINVLTGELLPTSGEVYTHENCR.I	YLR249W
target_0_64798_3_1	demo.ms2	2.87048	0.00220751	4.21444E-06	K.VINDAFGIEEGLMTTVHSLTATQK.T	YGR192C