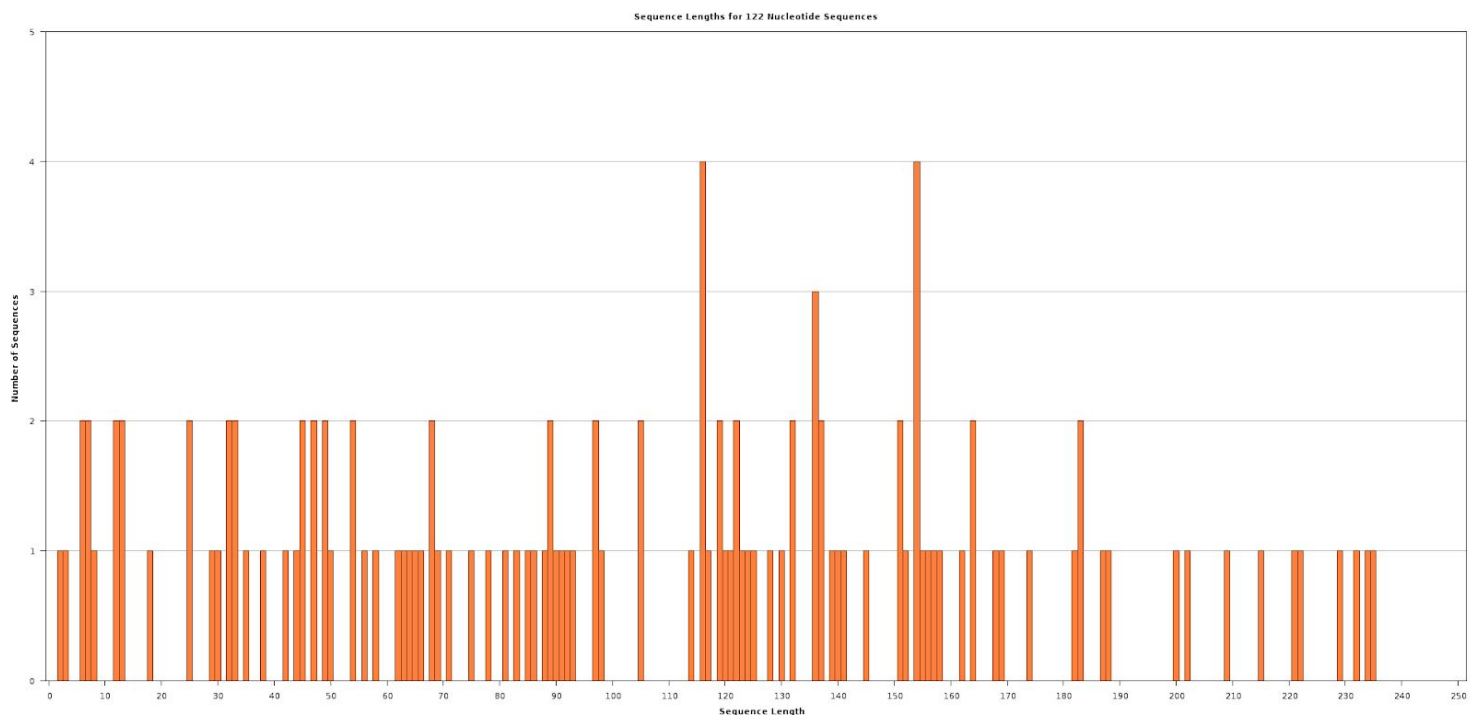


- **Species: Cinnamomum capparucoronae**
- **Data size: 1GB**
- **Comparison between ITSx output obtained using vsearch output as input and ITSxpress output obtained using contigs as input**

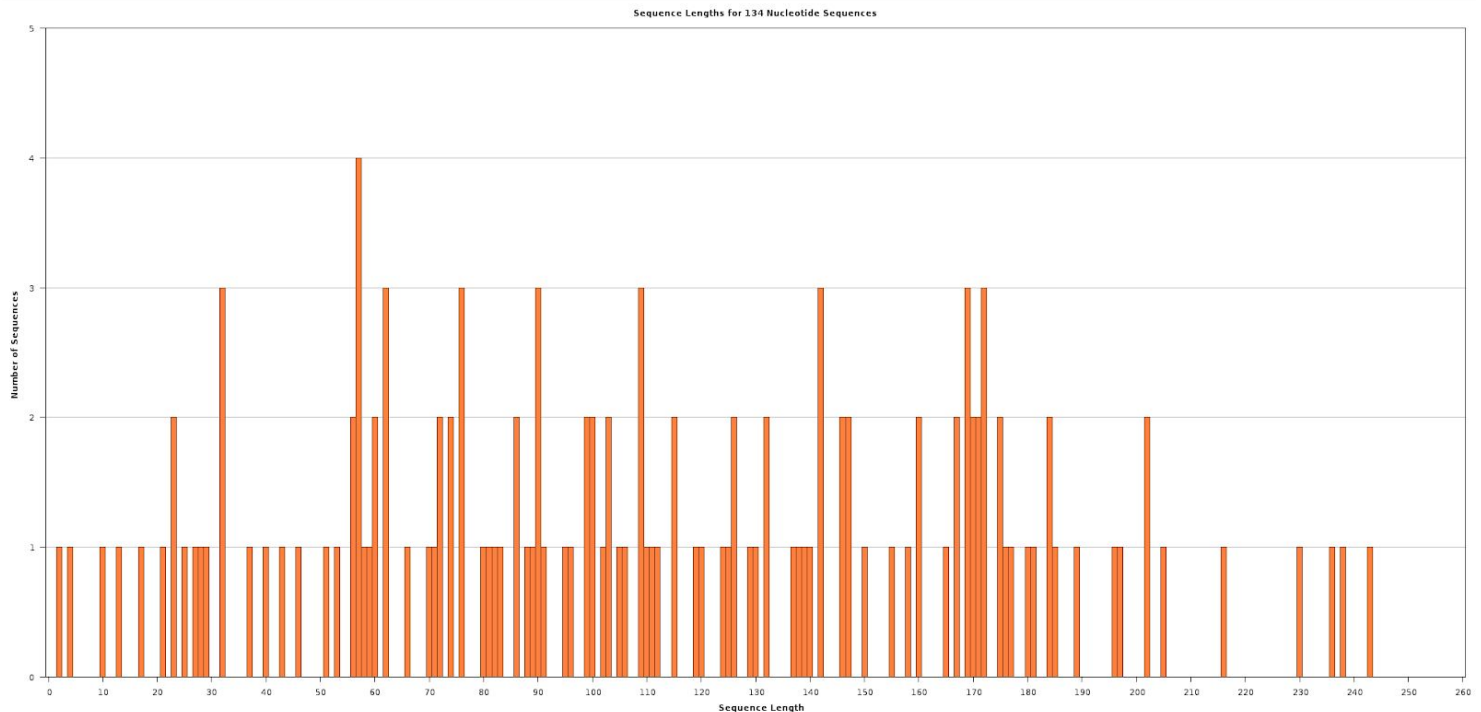
ITSx- Output

- creates separate sequence files for ITS1 and ITS2 (one FASTA file for the ITS1 and ITS2 regions)
- There are multiple sequences identified as candidate ITS1 and ITS2 sequences (considering all the merged reads)
- “ITSx is not designed to make accurate predictions on organism groups, and no double-checking of this prediction is performed. Therefore, the ITS sequences extracted should be further examined using e.g. BLAST searches.” - Taken from the ITSx user guide
- 122 candidate sequences for ITS2 and 134 candidate sequences for ITS1
- Length ranges of candidate sequences

❖ ITS2



❖ ITS1



Determine average lengths of ITS2 and ITS1 in plants

1. Yao, Hui & Song, Jingyuan & Liu, Chang & Luo, Kun & Jianping, Han & Li, Ying & Pang, Xiaohui & Xu, Hongxi & Zhu, Yingjie & Xiao, Pei-Gen & Chen, Shilin. (2010). Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. *PloS one*. 5. 10.1371/journal.pone.0013102.
 2. Liston, Aaron & Robinson, William & Oliphant, James & Alvarez-Buylla, Elena. (1996). Length Variation in the Nuclear Ribosomal DNA Internal Transcribed Spacer Region of Non-Flowering Seed Plants. *Systematic Botany*. 21. 109. 10.2307/2419742.
 3. Feng, Shangguo & Mengying, Jiang & Shi, Yujun & Jiao, Kaili & Shen, Chenjia & Lu, JiangJie & Ying, Qicai & Wang, Huizhong. (2016). Application of the Ribosomal DNA ITS2 Region of *Physalis* (Solanaceae): DNA Barcoding and Phylogenetic Study. *Frontiers in Plant Science*. 7. 10.3389/fpls.2016.01047.
 4. Wang, Xin-Cun & Liu, Chang & Huang, Liang & Bengtsson-Palme, Johan & Chen, Haimei & Zhang, Jian-Hui & Cai, Dayong & Li, Jian-Qin. (2015). ITS1: a DNA barcode better than ITS2 in eukaryotes?. *Molecular Ecology Resources*. 15. 573–586. 10.1111/1755-0998.12325.
- Considering the details in the above publications and Cinnamon is a dicot plant, the sequences with 100bp - 681bp and 101bp-522bp were taken as the candidate ITS1 and ITS2 regions respectively.
 - Accordingly, 62 sequences for ITS2 and 74 sequences for ITS1 were retained for further analyses.

Test1: Align all the sequences taken as ITS2 and ITS1 regions separately

ITS2

- The multiple alignment for the selected 62 sequences was done using the MAFFT algorithm
- Base-pair difference ranges from 0bp to 183bp
- The distance matrix was divided into 4 sections to have a deep evaluation

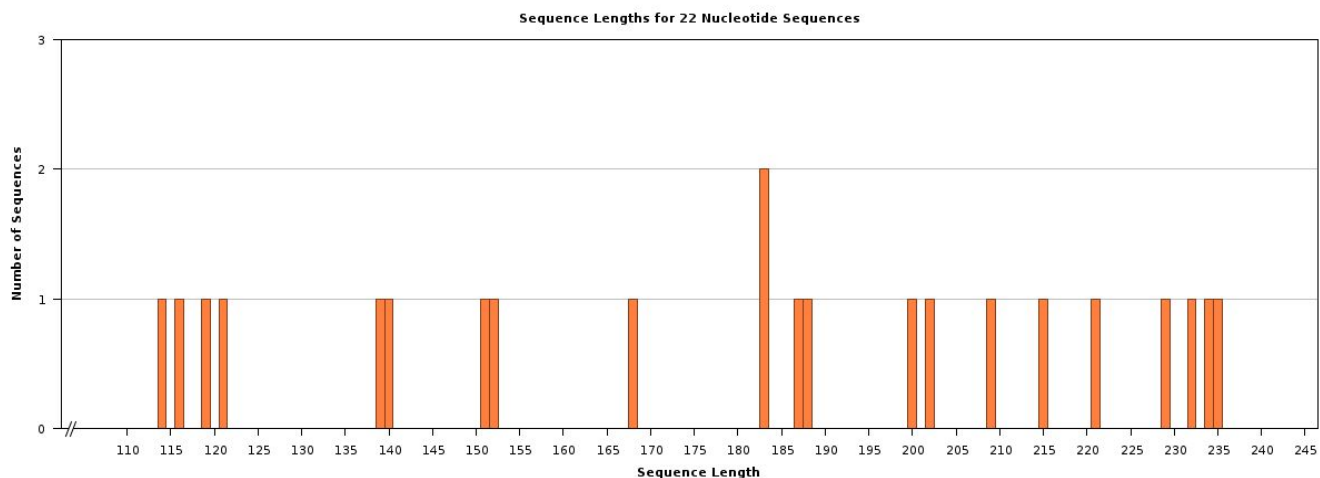
Section 1:

	E00526:96:HWIYCCXY:5:2105:25915:9255(T)ITS2:052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...	E0052...
E00526:96:HWIYCC...	0	0	0	0	0	0	1	1	0	3	0	0	0	2	2	1	1	1	1	3	2
E00526:96:HWIYCC...	0	0	0	0	0	0	1	1	0	3	0	0	0	2	2	1	1	1	1	3	2
E00526:96:HWIYCC...	0	0	0	0	0	0	1	1	0	3	0	0	0	2	2	1	1	1	1	3	2
E00526:96:HWIYCC...	0	0	0	0	0	0	1	1	0	3	0	0	0	2	2	1	1	1	1	3	2
E00526:96:HWIYCC...	0	0	0	0	0	0	1	1	0	3	0	0	0	1	1	1	1	1	1	3	2
E00526:96:HWIYCC...	0	0	0	0	0	0	1	1	0	3	0	0	0	1	1	1	1	1	1	3	2
E00526:96:HWIYCC...	1	1	1	1	1	1	2	0	0	4	0	0	0	3	3	2	1	1	1	3	3
E00526:96:HWIYCC...	1	1	1	1	1	1	2	1	1	2	1	1	1	3	3	2	2	2	2	4	3
E00526:96:HWIYCC...	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	1	1	1	3	2
E00526:96:HWIYCC...	3	3	3	3	3	3	4	2	1	1	1	1	5	5	4	3	2	2	2	4	5
E00526:96:HWIYCC...	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	1	1	1	3	2
E00526:96:HWIYCC...	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	1	1	1	3	2
E00526:96:HWIYCC...	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	1	1	1	3	2
E00526:96:HWIYCC...	2	2	2	2	1	1	3	3	1	5	1	1	1	1	0	0	0	0	0	2	1
E00526:96:HWIYCC...	2	2	2	2	1	1	3	3	1	5	1	1	1	0	0	0	0	0	0	2	1
E00526:96:HWIYCC...	1	1	1	1	1	1	2	2	1	4	1	1	1	0	0	0	0	0	0	2	1
E00526:96:HWIYCC...	1	1	1	1	1	1	2	2	1	3	1	1	1	0	0	0	0	0	0	2	1
E00526:96:HWIYCC...	1	1	1	1	1	1	2	1	2	2	1	1	1	0	0	0	0	0	0	2	1
E00526:96:HWIYCC...	1	1	1	1	1	1	2	1	2	1	1	1	1	0	0	0	0	0	0	2	1
E00526:96:HWIYCC...	1	1	1	1	1	1	2	1	2	1	1	1	1	0	0	0	0	0	0	2	1
E00526:96:HWIYCC...	1	1	1	1	1	1	1	2	1	2	1	1	1	0	0	0	0	0	0	2	1
E00526:96:HWIYCC...	1	1	1	1	1	1	1	2	1	2	1	1	1	0	0	0	0	0	0	2	1
E00526:96:HWIYCC...	3	3	3	3	3	3	3	4	3	4	3	3	3	2	2	2	2	2	2	4	3
E00526:96:HWIYCC...	2	2	2	2	2	2	3	3	3	2	5	2	2	2	1	1	1	1	1	3	

- This section contains the most similar sequences - 22 sequences
- Randomly selected sequences were blasted against the NCBI nr/nt database (02/16/2021; 1.30 pm - 2.00 pm)
- The blast results contained ITS2 regions of *Cinnamomum wightii*, *Cinnamomum chemungianum*, *Cinnamomum perrottetii*, *Cinnamomum travancoricum*, *Cinnamomum mathewianum*, *Cinnamomum dubium*, *Cinnamomum tamala*, *Cinnamomum litseifolium*, *Cinnamomum keralaense*, *Cinnamomum verum*, *Cinnamomum sulphuratum*, *Cinnamomum goaense*, *Cinnamomum agasthyamalayanum*, *Cinnamomum verum*, *Cinnamomum macrocarpum*.
- Thus, these sequences can be considered for further evaluation.

GC content : 73.0% - 76.0%

Length distribution



Section 2:

E00526:96:HWIYCC...	29	129	126	121	112	130	130	96	127	84	84	80	131	131	113	106	91	90	82	98	115
E00526:96:HWIYCC...	14	94	94	94	94	94	94	91	95	78	78	75	95	95	95	95	86	85	77	93	96
E00526:96:HWIYCC...	76	174	156	144	141	164	176	120	160	101	100	97	177	169	141	133	113	112	99	120	145
E00526:96:HWIYCC...	31	131	131	131	131	131	131	120	132	101	100	97	132	132	132	132	113	112	99	120	133
E00526:96:HWIYCC...	21	121	121	121	121	121	121	121	121	102	101	98	122	122	122	122	114	113	100	121	123
E00526:96:HWIYCC...	65	165	152	140	138	160	166	118	157	99	98	95	166	165	138	128	111	110	97	118	141
E00526:96:HWIYCC...	65	165	152	140	138	160	166	118	157	99	98	95	166	165	138	128	111	110	97	118	141
E00526:96:HWIYCC...	40	140	140	140	138	141	141	118	143	99	98	95	141	141	138	128	111	110	97	118	141
E00526:96:HWIYCC...	31	131	131	131	131	131	132	118	133	99	98	95	132	132	132	128	111	110	97	118	133
E00526:96:HWIYCC...	30	130	130	130	130	130	131	118	132	99	98	95	131	131	131	128	111	110	97	118	132
E00526:96:HWIYCC...	57	157	145	133	130	153	157	111	147	93	92	90	158	157	130	121	104	103	92	113	134
E00526:96:HWIYCC...	17	117	117	117	117	117	117	111	118	93	92	90	118	118	118	118	104	103	92	113	119
E00526:96:HWIYCC...	58	158	146	134	130	153	158	110	148	92	91	89	159	158	130	121	103	102	91	112	134
E00526:96:HWIYCC...	60	160	147	135	131	155	160	111	149	93	92	90	161	160	131	122	104	103	92	113	135
E00526:96:HWIYCC...	33	133	133	133	131	134	133	112	133	94	93	91	134	134	131	122	105	104	93	114	135
E00526:96:HWIYCC...	27	127	127	127	127	128	127	112	128	93	92	90	128	128	128	122	105	104	92	114	129
E00526:96:HWIYCC...	14	114	114	114	114	114	114	110	115	92	91	89	115	115	115	115	104	103	91	112	116
E00526:96:HWIYCC...	47	147	143	133	129	148	148	114	147	94	92	88	148	148	130	121	107	106	90	116	133
E00526:96:HWIYCC...	29	129	129	129	129	130	130	114	131	94	92	88	130	130	130	121	107	106	90	116	131
E00526:96:HWIYCC...	15	115	115	115	115	115	116	114	117	94	92	88	116	116	116	116	107	106	90	116	117
E00526:96:HWIYCC...	56	156	152	140	135	157	156	115	154	94	92	89	157	157	135	124	108	107	91	117	139
E00526:96:HWIYCC...	24	124	124	124	124	124	124	119	125	99	97	94	125	125	125	125	114	113	96	121	126
E00526:96:HWIYCC...	65	165	159	150	146	166	165	120	159	97	96	92	167	167	146	135	113	112	94	122	149
E00526:96:HWIYCC...	38	138	138	136	133	139	138	109	139	93	93	89	139	139	133	121	105	104	91	111	136
E00526:96:HWIYCC...	35	135	135	135	135	136	135	109	135	92	90	87	136	136	135	124	105	104	89	111	137
E00526:96:HWIYCC...	12	112	112	112	112	112	113	103	114	89	89	85	113	113	113	113	98	97	87	106	114
E00526:96:HWIYCC...	27	127	127	127	127	127	128	122	127	99	98	94	128	128	128	128	113	112	96	122	129
E00526:96:HWIYCC...	82	182	182	182	182	182	182	178	183	135	134	130	183	183	183	183	147	146	132	178	184
E00526:96:HWIYCC...	21	121	121	121	121	121	121	121	121	99	97	92	122	122	122	122	118	117	94	122	123
E00526:96:HWIYCC...	44	144	144	144	144	144	144	135	145	111	109	108	145	145	145	145	125	124	109	135	146
E00526:96:HWIYCC...	24	124	124	124	124	124	124	124	125	100	98	96	125	125	125	125	116	115	97	124	126
E00526:96:HWIYCC...	25	125	125	125	125	125	125	125	126	101	99	97	126	126	126	126	117	116	98	125	127
E00526:96:HWIYCC...	09	109	109	109	109	109	109	109	110	85	83	81	109	109	109	109	100	99	81	108	109
E00526:96:HWIYCC...	05	105	105	105	105	105	105	105	106	81	79	77	105	105	105	105	96	95	77	104	105
E00526:96:HWIYCC...	00	100	100	100	100	100	100	100	101	76	74	72	100	100	100	100	91	90	72	99	100
E00526:96:HWIYCC...	08	98	98	98	98	98	98	98	99	74	72	70	98	98	98	98	89	88	70	97	98
E00526:96:HWIYCC...	17	97	97	97	97	97	97	97	97	73	71	69	97	97	97	97	88	87	69	96	97
E00526:96:HWIYCC...	19	89	89	89	89	89	89	89	89	65	63	61	89	89	89	89	80	79	61	89	89
E00526:96:HWIYCC...	04	104	104	104	104	104	104	104	105	80	78	76	104	104	104	104	95	94	76	103	104
E00526:96:HWIYCC...	03	103	103	103	103	103	103	103	104	79	77	75	103	103	103	103	94	93	75	102	103

E00526:96:HWIYCC...	108	109	125	125	113	116	116	116	116	118	114	114	116	115	117	110	112	112	108	118	119	120	110	105	112	184	114
E00526:96:HWIYCC...	85	90	104	104	96	101	101	101	101	101	90	92	91	90	92	94	94	90	90	90	90	93	103	105	101	112	179
E00526:96:HWIYCC...	169	169	161	161	155	162	162	162	162	168	170	167	169	170	170	172	177	177	177	175	170	169	162	169	184	176	
E00526:96:HWIYCC...	86	86	86	86	87	88	88	88	88	90	90	88	90	91	90	90	87	87	87	87	94	84	86	80	86	171	
E00526:96:HWIYCC...	130	129	129	129	120	127	127	127	127	127	126	125	125	123	123	126	123	133	133	131	133	132	132	134	131	131	
E00526:96:HWIYCC...	129	125	126	126	125	124	124	124	124	124	122	123	124	123	123	125	124	133	133	133	133	125	128	131	128	123	
E00526:96:HWIYCC...	130	126	125	125	124	123	123	123	123	123	123	124	125	124	124	126	125	133	133	133	133	125	128	132	129	124	
E00526:96:HWIYCC...	114	110	109	109	108	107	107	107	107	106	107	108	107	107	109	108	116	116	116	108	111	114	112	107	115		
E00526:96:HWIYCC...	110	106	105	105	104	103	103	103	103	103	102	103	104	103	103	105	104	112	112	112	104	107	110	108	103	113	
E00526:96:HWIYCC...	104	100	99	99	98	97	97	97	97	96	97	98	97	97	99	98	106	106	106	98	101	104	102	97	108		
E00526:96:HWIYCC...	102	97	97	97	96	95	95	95	95	95	94	95	96	95	94	96	96	102	102	102	97	99	100	98	93		
E00526:96:HWIYCC...	101	96	96	96	95	94	94	94	94	94	93	94	95	94	93	95	95	101	101	101	101	96	98	99	97		
E00526:96:HWIYCC...	93	88	90	90	89	88	88	88	88	87	87	88	87	86	88	88	93	93	93	88	90	92	90	85	96		
E00526:96:HWIYCC...	107	109	103	103	102	99	99	99	99	99	102	103	104	103	103	105	104	108	108	108	103	104	106	109	103	110	
E00526:96:HWIYCC...	106	101	102	102	101	98	98	98	98	98	101	102	103	102	102	104	103	107	107	107	102	103	105	108	102	109	

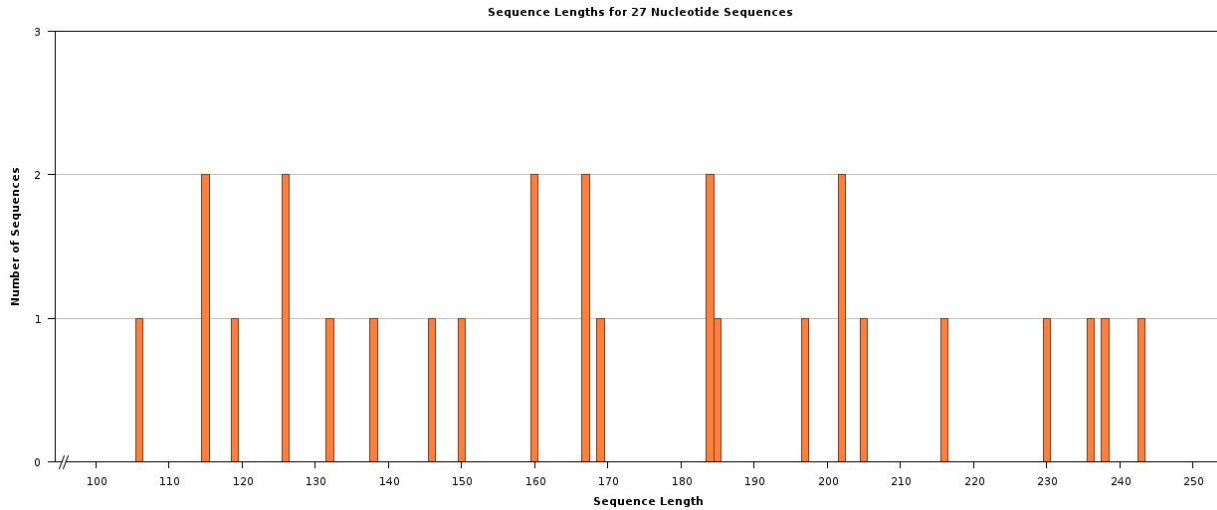
- Randomly selected sequences were blasted against the NCBI nr/nt database (02/16/2021; 2.40 pm - 2.55 pm)
- The blast results contained ITS2 regions of fungi species (*Curvibasidium pallidicorallinum*, Basidiomycetes, *Colletotrichum nymphaeae*, *Colletotrichum scovillei*, *Elsinoaceae*, *Myriangiales*, *Diaporthe velutina*, *Diaporthe inconspicua*, *Rhodotorula*).

Section 3:

E00529.98:HWYCC00	38	122	95	74	121	121	103	89	116	75	115	113	85	85	75	105	81	66	107	71	127	103	69	108	95		
E00529.98:HWYCC00	38	73	73	69	76	76	76	76	76	76	66	71	69	66	69	64	64	64	63	68	74	59	64	109	90		
E00529.98:HWYCC00	122	73	0	0	3	35	35	30	27	26	69	51	66	65	63	63	50	93	77	65	90	63	109	96	125	104	
E00529.98:HWYCC00	85	73	0	0	3	23	23	23	23	23	58	51	54	54	58	59	50	70	70	65	71	63	81	78	78	125	104
E00529.98:HWYCC00	74	69	3	3	13	13	13	13	13	46	41	42	42	46	46	44	58	58	58	60	55	70	69	66	113	96	
E00529.98:HWYCC00	121	76	35	23	13	0	0	0	0	65	48	62	62	61	59	50	95	78	63	86	59	111	95	92	116	101	
E00529.98:HWYCC00	121	76	35	23	13	0	0	0	0	65	48	62	62	61	59	50	95	78	63	86	59	111	95	92	116	101	
E00529.98:HWYCC00	103	76	30	23	13	0	0	0	0	60	48	57	58	61	59	50	80	78	63	84	59	91	92	116	101		
E00529.98:HWYCC00	86	76	27	13	0	0	0	0	0	56	48	53	53	57	57	50	73	72	63	77	59	87	81	85	118	100	
E00529.98:HWYCC00	85	76	26	23	13	0	0	0	0	56	48	52	53	57	57	50	72	72	63	76	59	86	80	84	116	101	
E00529.98:HWYCC00	116	72	69	58	46	65	65	60	56	56	15	15	15	20	18	15	20	82	66	55	70	63	103	95	91	118	90
E00529.98:HWYCC00	75	66	51	51	41	48	48	48	48	15	15	14	13	16	12	52	52	50	59	63	73	73	70	114	92		
E00529.98:HWYCC00	115	71	66	54	42	62	62	57	52	52	15	15	9	12	13	16	79	62	51	72	62	101	96	90	114	91	
E00529.98:HWYCC00	113	69	65	54	42	62	62	61	53	53	20	14	9	5	15	15	77	60	49	71	61	100	94	90	116	90	
E00529.98:HWYCC00	85	66	63	58	46	61	61	57	57	18	13	12	5	10	16	60	58	47	67	61	84	84	88	115	92		
E00529.98:HWYCC00	69	66	63	59	46	59	59	58	57	15	15	16	13	15	10	13	60	60	57	63	78	79	78	117	90		
E00529.98:HWYCC00	75	69	50	50	44	50	50	50	50	50	12	16	15	16	13	51	51	51	60	63	73	70	65	110	94		
E00529.98:HWYCC00	105	64	93	70	58	95	95	80	73	72	82	52	79	77	60	60	51	0	0	0	69	47	95	84	77	112	90
E00529.98:HWYCC00	81	64	77	70	58	78	78	78	73	72	66	52	62	60	58	60	51	0	0	0	51	47	78	75	75	112	90
E00529.98:HWYCC00	66	64	65	65	58	63	63	63	63	63	55	50	51	49	47	52	51	0	0	0	42	45	66	62	60	108	90
E00529.98:HWYCC00	107	63	80	71	60	86	86	84	77	76	70	69	72	71	67	67	60	69	51	42	31	113	96	90	118	90	
E00529.98:HWYCC00	107	63	80	71	60	86	86	84	77	76	70	69	72	71	67	64	63	47	47	45	31	83	76	74	119	93	
E00529.98:HWYCC00	127	74	106	81	113	97	97	103	73	101	100	84	78	85	78	66	113	83	58	62	58	120	97	120	105	105	
E00529.98:HWYCC00	103	59	96	78	89	95	95	91	81	80	95	73	96	94	84	79	76	84	73	62	96	76	58	46	110	105	
E00529.98:HWYCC00	89	64	89	78	66	92	92	82	83	84	91	70	90	90	88	82	77	73	60	74	67	46	105	101	105	105	

Section 4:

- Randomly selected sequences were blasted against the NCBI nr/nt database (02/16/2021; 3.14 pm - 3.20 pm)
- The blast results contained ITS2 regions of fungi species (*Colletotrichum acutatum*, *Colletotrichum nymphaeae*, *Colletotrichum fructicola*, *Colletotrichum gloeosporioides*, *Colletotrichum fragariae*, *Colletotrichum siamense*).

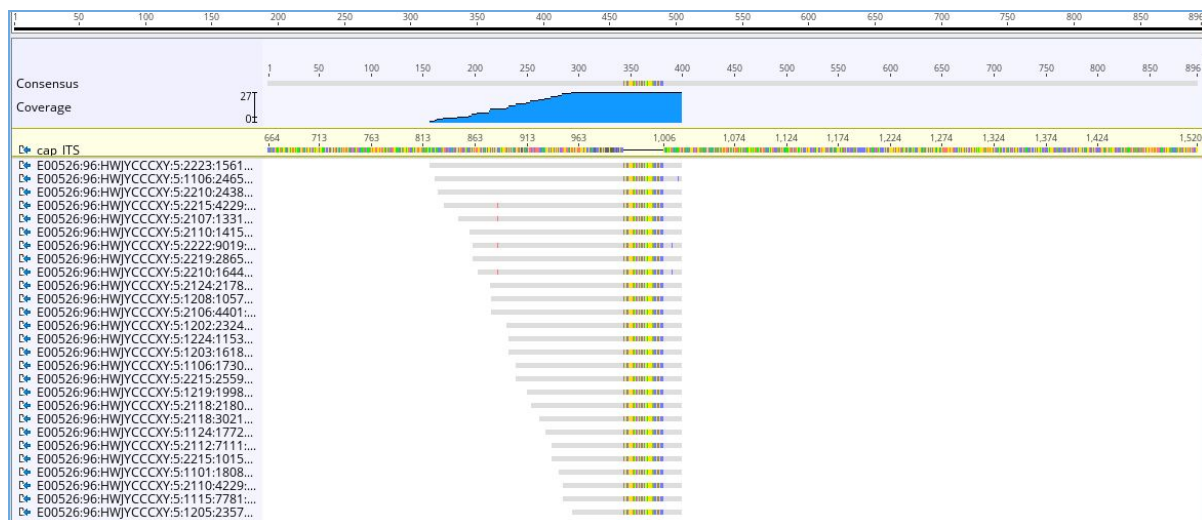


Section 2:

- Randomly selected sequences were blasted against the NCBI nr/nt database (02/16/2021; 4.05 pm - 4.20 pm)
- The blast results contained ITS2 regions of fungi species (*Ceramothyrium longivolcaniforme*, *Cyphellophora pruni*, *Neofusicoccum parvum*, *Diaporthe guangxiensis*, *Diaporthe searlei*, *Diaporthe lithocarpus*, *Phomopsis phyllanthicola*, *Colletotrichum scovillei*, *Colletotrichum fioriniae*).
- Thus, these sequences can be discarded from the analysis.

❖ **Test1.2: All these 74 sequences were mapped to the *Cin. capparu-coronde* ITS (ITS1-5.8s-ITS2-26s) sequence which I (sandali) constructed.**

❖ **Interestingly, only 27 sequences in Section 1 were mapped.**

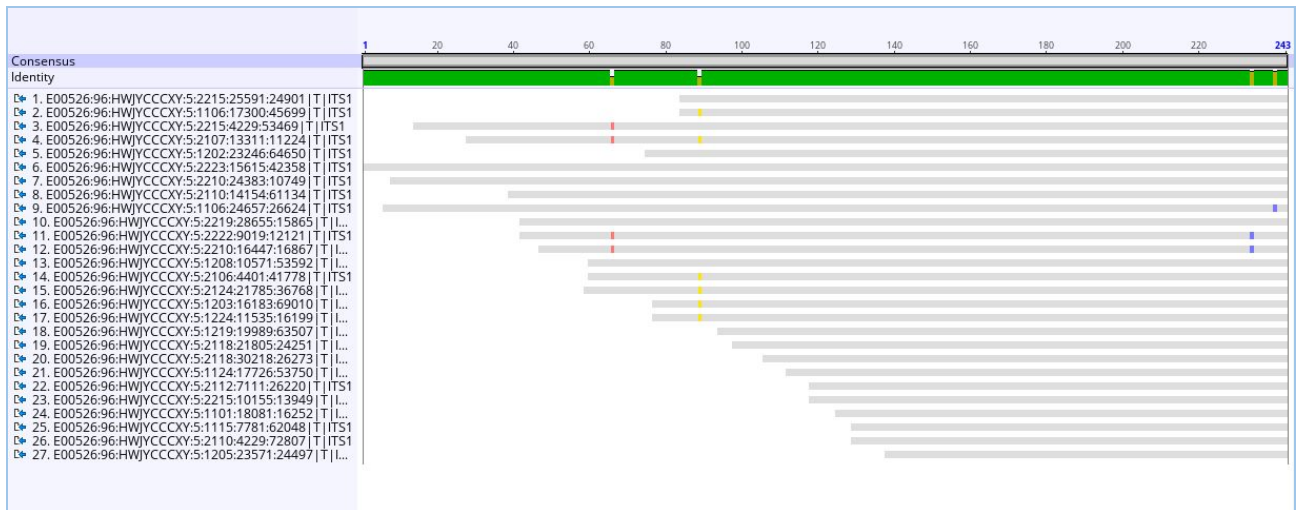


- Since there are 22 sequences for ITS2 and 27 for ITS1, the consensus sequences taken from the multiple alignments are considered to compare ITSx and ITSxpress results. (Taking consensus doesn't negatively impact on further comparisons- there are a maximum of 5 base-pair difference

ITS2 alignment view



ITS1 alignment view



Test2: Compare the results from ITSx and ITSxpress

- ITS1 (consensus - ITSX_ITS1) and ITS2 (consensus - ITSX_ITS2) sequences taken from ITSx and the output sequence of ITSxpress (itsxpress) were aligned using the MAFFT algorithm.

Alignment view



Distance matrix

	itsxpress	ITSX ITS1	ITSX ITS2
itsxpress		0	0
ITSX ITS1	0		0
ITSX ITS2	0	0	

- According to the above results, there is no difference in the results taken from ITSx and ITSxpress