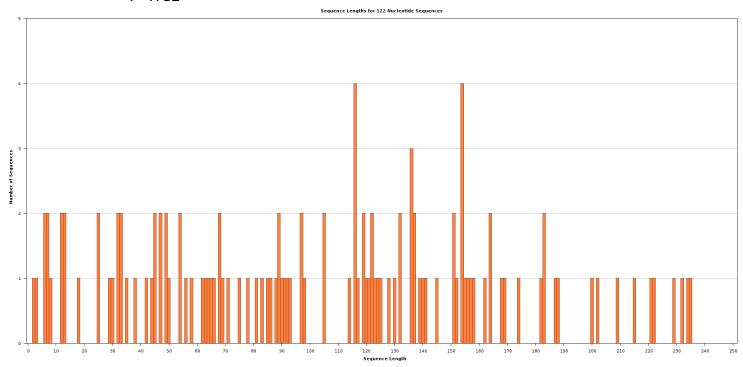
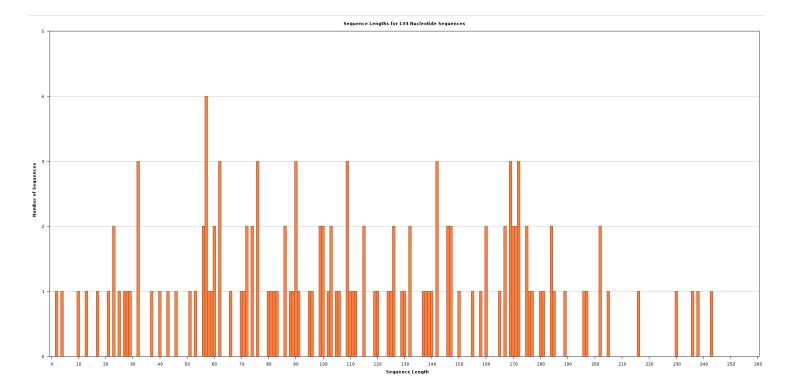
- Species: Cinnamomum capparu-coronde
- Data size: 1GB
- Comparison between ITSx output obtained using vsearch output as input and ITSxpress output obtained using contigs as input

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





Determine average lengths of ITS2 and ITS1 in plants

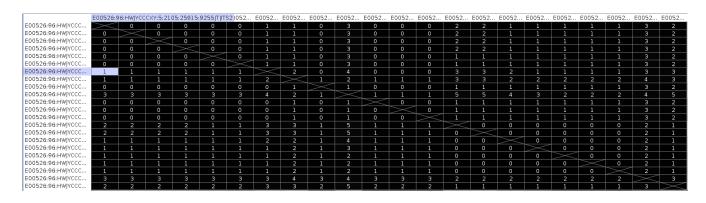
- 1. Yao, Hui & Song, Jingyuan & Liu, Chang & Luo, Kun & Jianping, Han & Li, Ying & Pang, Xiaohui & Xu, Hongxi & Zhu, Yingjie & Xiao, Pei-Gen & Chen, Shilin. (2010). Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. PloS one. 5. 10.1371/journal.pone.0013102.
- 2. Liston, Aaron & 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.
- 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

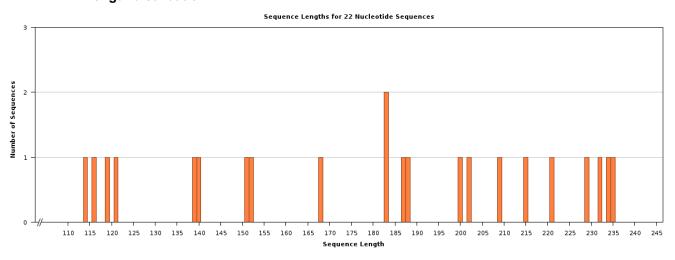
- The multiple alignment for the selected 62 sequences was done using the MAFFT algorithm
- Base-pair difference ranges from 0bp to 183bp
- The distance matrix was divided into 4 sections to have a deep evaluation

Section 1:



- 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



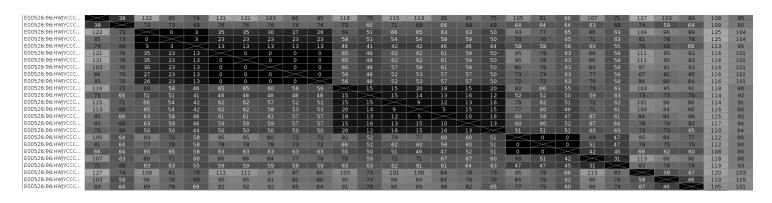
Section 2:

IE00526:96:HWIYCCC... 25

E00526:96			12	9	126	121	112	130	130	96	127	84	84	80	131	131	113	106	91			82	98	115	
E00526:96	5:HWJYCC	C 14	9	4	94	94	94	94	94	91	95	78	78	75	95	95	95	95	86	85	5	77	93	96	
E00526:96	5:HWJYCC	C 76	17	4	156	144	141	164	176	120	160	101	100	97	177	169	141	133	113	3 11	2	99	120	145	
E00526:96	B:HWJYCC	C 31	13	1	131	131	131	131	131	120	132	101	100	97	132	132	132	132	113	11	2	99	120	133	
E00526:96	5:HWJYCC	C 21	12	1	121	121	121	121	121	121	121	102	101	98	122	122	122	122	114	11	3	100	121	123	1
E00526:96	S:HWJYCC	C 65	16	5	152	140	138	160	166	118	157	99	98	95	166	165	138	128	111	. 11	0	97	118	141	
E00526:96	3:HWJYCC	C 65	16	5	152	140	138	160	166	118	157	99	98	95	166	165	138	128	111	. 11	0	97	118	141	I
E00526:96	s:HWJYCC	C 40	14	0	140	140	138	141	141	118	143	99	98	95	141	141	138	128	111	. 11	0	97	118	141	
E00526:96	B:HWIYCC	C 31	13	1	131	131	131	131	132	118	133	99	98	95	132	132	132	128	111	. 11	0	97	118	133	1
E00526:96	B:HWJYCC	С 30	13		130	130	130	130	131	118	132	99	98	95	131	131	131		111			97	118	132	
E00526:96	s:HWIYCC	C 57	15	7	145	133	130	153	157	111	147	93	92	90	158	157	130	121	104	1 10	3	92	113	134	
E00526:96					117	117	117	117	117	111	118	93	92	90	118	118	118		104			92	113	119	1
E00526:96	B:HWIYCC	C 58			146	134	130	153	158	110	148	92	91	89	159	158	130		103			91	112	134	1
E00526:96					147	135	131	155	160	111	149	93	92	90	161	160	131					92	113	135	
E00526:96			13		133	133	131	134	133	112	133	94	93	91	134	134	131					93	114	135	1
E00526:96			12		127	127	127	128	127	112	128	93	92	90	128	128	128					92	114	129	
E00526:96					114	114	114	114	114	110	115	92	91	89	115	115	115		104			91	112	116	
E00526:96					143	133	129	148	148	114	147	94	92	88	148	148	130		107			90	116	133	
E00526:96			12		129	129	129	130	130	114	131	94	92	88	130	130	130		107			90	116	131	1
E00526:96					115	115	115	115	116	114	117	94	92	88	116	116	116		107			90	116	117	1
E00526:96					152	140	135	157	156	115	154	94	92	89	157	157	135		108			91	117	139	
E00526:96					124	124	124	124	124	119	125	99	97	94	125	125	125		114			96	121	126	
E00526:96					159	150	146	166	165	120	159	99	96	92	167	167	146		113			94	122	149	
E00526:96					138	136	133	139	138	109	139	97	98	89	139	139	133		105			91	111	136	
E00526:96			13		135	135	135	136	135	109	135	93	90	87	139	136	135		105			89	111	137	
E00526:96					112	112	112	112	113	103	114	89	89	85	113	113	1133		98			87	106	114	
E00526:96			12		127	127	127	127	128	122	127	99	98	94	128	128	128		113			96	122	129	
E00526:96					182	182	182	182	182	178	183	135	134		183	183	183					132	178	184	1
E00526:96			12		121	121	121	121	121	121	121	99	97	92	122	122	122		118			94	122	123	i
E00526:96					144	144	144	144	144	135	145	111	109		145	145	145					109	135	146	1
E00526:96					124	124	124	124	124	124	125	100	98	96	125	125	125		116			97	124	126	
E00526:96			12		125	125	125	125	125	125	126	101	99	97	126	126	126					98	125	127	
E00526:96			10		109	109	109	109	109	109	110	85	83	81	109	109	109		100			81	108	109	
E00526:96			10		105	105	105	105	105	105	106	81	79	77	105	105	105		96			77	104	105	
E00526:96			10		100	100	100	100	100	100	100	76	74		100	100	100					72	99	100	
E00526:96			9:		98	98	98	98	98	98	99	74	72	70	98	98	98	98	89			70	97	98	
E00526:96			9		97	97	97	97	97	97	98	73	71	69	97	97	97	97	88			69	96	97	
E00526:96			8		89	89	89	89	89	89	90	65	63	61	89	89	89	89	80			61	89	89	
E00526:96			10		104	104	104	104	104	104	105	80	78	76	104	104	104		95			76	103	104	
E00526:96					103	103	103	103	103	103	103	79	77	75	103	103	103					75	102	103	
1		_																							101
E00526:96:HWJYCCC E00526:96:HWJYCCC	108 85	109 90	125 104	125	113 96	116	116	116 11 101 10		118 90	114 92	114 91		115 117 92 94	110 94	112 90	112	108 11 90 9			110			112	184 114 179 103
E00526:96:HWJYCCC	169	169	161	161	155	162	162	162 16		168	170			170 170			177	177 17			162			179	171
E00526:96:HWJYCCC	86	85	86	86	87	88	88	88 8	88	90	90	88		91 90	90	87	87	87 9		86	90	86	114	103	171
E00526:96:HWJYCCC	138	129	129	129	120	127	127	127 12		126	125	125		123 126				131 13			134			116	186 128
E00526:96:HWJYCCC	129	125	126	126	125	124	124	124 12		122	123	124		123 125		133		133 12			128			125	190 125
E00526:96:HWJYCCC E00526:96:HWJYCCC	130 114	126 110	125 109	125	124	123	123	123 12 107 10		123 106	124	125 108		124 126 107 109		133 116		133 12 116 10			129			123 107	190 125 173 111
E00526:96:HWJYCCC	110	106	105	105	108	107	107	107 10		100	107	108		103 105	108			112 10			108			107	171 107
E00526:96:HWJYCCC	104	100	99	99	98	97	97	97 9		96	97	98		97 99	98			106 9			102			97	165 101
E00526:96:HWJYCCC	102	97	97	97	96	95	95	95 9		94	95	96		94 96	96			102 9		100	98			96	162 98
E00526:96:HWJYCCC	101	96	96	96	95	94	94	94 9		93	94	95		93 95	95	101		101 9		99	97	92		95	161 97
E00526:96:HWJYCCC E00526:96:HWJYCCC	93 107	88 102	90 103	90 103	89 102	99	99	99 9		87 102	103	88 104		86 88 103 105	104	93 108	93 108	93 8 108 10		92 106	90 109			87 100	148 92 170 102
E00526:96:HWJYCCC	107	102	103	103	102	99	98	98 9		102	103			103 105				108 10			109			99	169 101 -
	_			-																		_			

- 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).

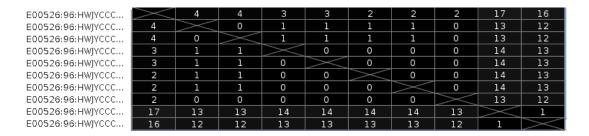
Section 3:



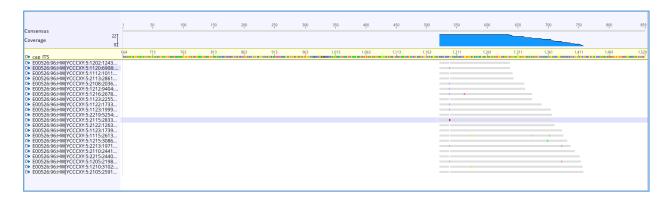
Randomly selected sequences were blasted against the NCBI nr/nt database (02/16/2021; 3.00 pm - 3.10 pm)

 The blast results contained ITS2 regions of fungi species (Neopestalotiopsis saprophytica, Neopestalotiopsis formicarum, Neopestalotiopsis foedans, Pestalotiopsis scorpina, Pestalotiopsis maculans, Diaporthe searlei, Diaporthe drenthii, Neofusicoccum parvum, Neofusicoccum yunnanense).

Section 4:



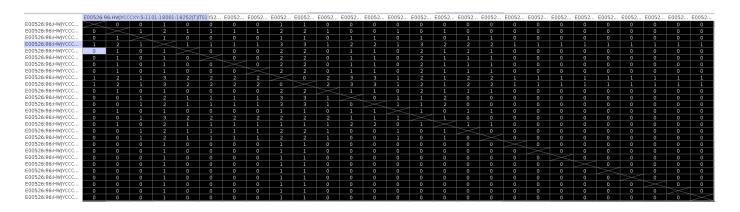
- Randomly selected sequences were blasted against the NCBI nr/nt database (02/16/2021; 3.14 pm 3.20 pm)
- The blast results contained ITS2 regions of fungi species (Colletotrichum acutatum, Colletotrichum nymphaeae, Colletotrichum fructicola, Colletotrichum gloeosporioides, Colletotrichum fragariae, Colletotrichum siamense).
- Considering the above results, sequences considered in sections 2-4 can be discarded from the analysis.
- **❖** Test1.1: All these 62 sequences were mapped to the Cin. capparu-coronde ITS (ITS1-5.8s-ITS2-26s) sequence which I (sandali) constructed.
- Interestingly, only 22 sequences in Section 1 were mapped.



ITS1

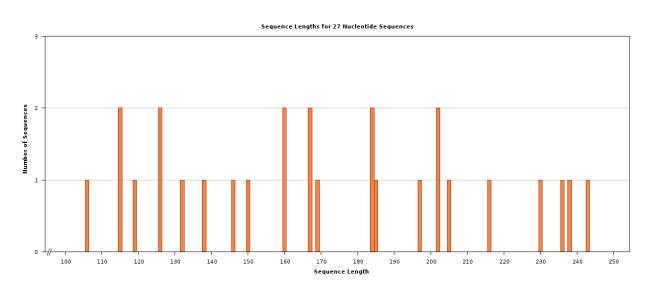
- Same as above, all the 74 sequences were aligned using the MAFFT algorithm and the base-pair difference in the resulting distance matrix was 0bp -196bp.
- The read alignment was evaluated by dividing the distance matrix into two sections.

Section 1



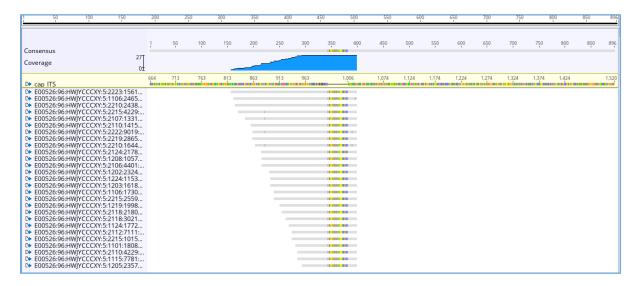
- This section contains the most similar sequences 27 sequences
- Randomly selected sequences were blasted against the NCBI nr/nt database (02/16/2021; 3.47 pm 4.00 pm)
- The blast results contained ITS2 regions of Cinnamomum tonkinense, Cinnamomum verum, Cinnamomum multiflorum, Cinnamomum iners, Cinnamomum dictyoneuron, Cinnamomum bejolghota, Cinnamomum kotoense, Cinnamomum rhynchophyllum, Cinnamomum javanicum, Cinnamomum cordatum, Cinnamomum tamala, Cinnamomum grandifolium)
- Thus, these sequences can be considered for further evaluation.

GC content : 71.3% - 75.5% Length distribution



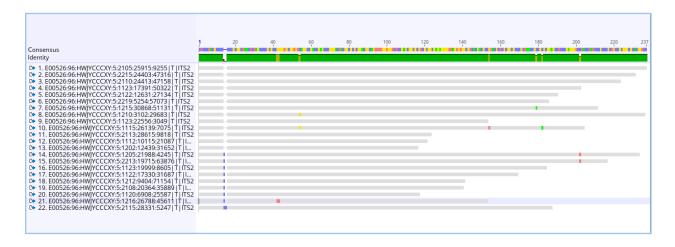
Section 2:

- 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.

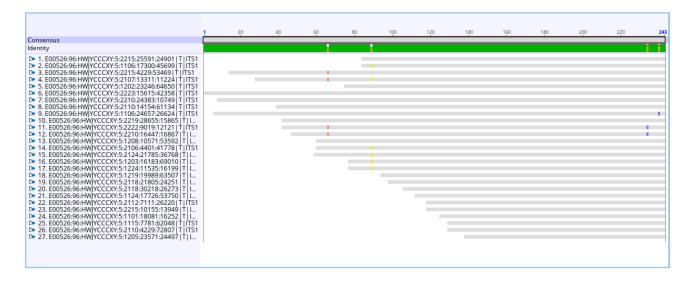


 Since there are 22 sequences for ITS2 and 27 for ITS1, the consensus sequences taken from the multiple alignments are considered to compare ITSx and ITSxpress results. ()Taking consensus doesn't negatively impact on further comparisonsthere are a maximum of 5 base-pair difference

ITS2 alignment view



ITS1 alignment view



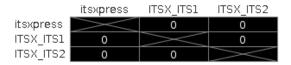
Test2: Compare the results from ITSx and ITSxpress

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



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