

Supplementary File

Differential Expressions of the *Hirsutella Sinensis* Genes and Intraspecific Genetic Variations among *H. Sinensis* Strains

Xiu-Zhang Li, Yu-Ling Li, Yi-Sang Yao, Jia-Shi Zhu*

Table S1: 312 strains used for PCR amplification of *H. sinensis* genes (in an alphabetical order).

1229	CUHK:CD8	GZUH2010RW4	ISO-29	QH09_281	SC-3	XZ09_59	XZ-NQ_84
ARSEF 6282	CUHK:CD15	GZUH2010RW5	ISO-30	QH10_1	SC-4	XZ09_71	XZ-NQ_86
CGB999335	CUHK:CDC1	Haplotype 1	ISO-31	QH10_4	SC-5	XZ09_80	XZ-NQ_92
Co18	CUHK:CDC2-1	Haplotype 2	ISO-32	QH10_7	SC-7	XZ09_95	XZ-NQ_139
CS09-111	CUHK:CDC2-2	Haplotype 3	ISO-33	QH-LJ-188	SCK05-4-3	XZ09_100	XZ-NQ_154
CS09-121	CUHK:CDC3	Haplotype 4	ISO-34	QH-LJ-189	XZ05_2	XZ09_106	XZ-NQ_155
CS09-143	CUHK:CDC4	Haplotype 5	ISO-35	QH-LJ-197A	XZ05_3	XZ09_113	XZ-NQ_165
CS09-225	CUHK:CDC5	Haplotype 6	ISO-36	QH-LJ-197B	XZ05_6	XZ09_118	XZ-NQ_166
CS09-229	CUHK:CDC6	Haplotype 7	ISO-37	QH-LJ-199	XZ05_7	XZ10_7	XZ-NQ_176
CS18-226	CUHK:CDC7	HMIGD 20940	ISO-38	QH-LJ-206	XZ05_8	XZ10_15	XZ-NQ_180
CS18-266	CUHK:CRE2	ID10_1	ISO-39	QH-LJ-207	XZ05_12	XZ10_17	XZ-SN_44
CS2	CUHK:CRE3	IOZ07	ISO-40	QH-LJ-214	XZ06_44	XZ10_23	YHH 1805
CS25-273	CUHK:CRE4	ISO-1	NP10_1	QH-LJ-236	XZ06_124	XZ12_1	YN07_6
CS26-277	CUHK:CSC1	ISO-2	NP10_2	QH-YS-196	XZ06_152	XZ12_16	YN07_8
CS34-291	CUHK:CSC2	ISO-3	QH06_197	QH-YS-199	XZ06_260	XZ12_33	YN09_3
CS36-1294	CUHK:CSC3	ISO-4	QH07_188	SC09_1	XZ07_108	XZ12_43	YN09_6
CS37-295	CUHK:CXB1	ISO-5	QH07_197	SC09_21	XZ07_11	XZ-CD_4	YN09_22
CS526-277	CUHK:CXB2	ISO-6	QH09_11	SC09_36	XZ07_46	XZ-CD_26	YN09_51
CS560-961	CUHK:CXB4	ISO-7	QH09_20L	SC09_37	XZ07_133	XZ-CD_30	YN09_61
CS561-964	CUHK:CXB5	ISO-8	QH09_22L	SC09_47	XZ07_154	XZ-CD_41	YN09_64
CS6-251	CUHK:CXB6	ISO-9	QH09_33L	SC09_57	XZ07_166	XZ-CD_59	YN09_72
CS68-2-1228	DCXC8	ISO-10	QH09_37	SC09_65	XZ07_176	XZ-CD_64	YN09_81
CS68-2-1229	EFCC7287	ISO-11	QH09_46	SC09_77	XZ07_180	XZ-CD_A1	YN09_85
CS68-5-1216	GYOKUJU	ISO-12	QH09_56	SC09_87	XZ07_H1	XZ-CD_B1	YN09_89
CS70-1208	GS09_111	ISO-13	QH09_66	SC09_88	XZ07_H2	XZ-CD7_10	YN09_96
CS70-1211	GS09_121	ISO-14	QH09_78	SC09_89	XZ08_4	XZ-LZ05_6	YN09_101
CS70-1212	GS09_131	ISO-15	QH09_93	SC09_97	XZ08_10	XZ-LZ06_1	YN09_140
CS71-1218	GS09_143	ISO-16	QH09_111	SC09_107	XZ08_24	XZ-LZ06_7	YN-1

CS71-1219	GS09_201	ISO-17	QH09_122	SC09_117	XZ08_26	XZ-LZ06_21	YN-4
CS71-1220	GS09_225	ISO-18	QH09_131	SC09_128	XZ08_33	XZ-LZ06_61	YN-5
CS76-1284	GS09_229	ISO-19	QH09_143	SC09_147	XZ08_38	XZ-LZ06_108	YN-6
CS91-1291	GS09_255	ISO-20	QH09_151	SC09_157	XZ08_56	XZ-LZ07_108	YN-8
CUHK: CBD1	GS09_281	ISO-21	QH09_157	SC09_167	XZ08_59	XZ-LZ07_30	ZJB12195
CUHK: CD1	GS09_288	ISO-22	QH09_164	SC09_180	XZ08_A1	XZ-LZ07_64	
CUHK: CD2	GS09_311	ISO-23	QH09_173	SC09_190	XZ08_B1	XZ-LZ07_H1	
CUHK: CD3	GS09_337	ISO-24	QH09_187	SC09_200	XZ09_4	XZ-LZ07_H2	
CUHK: CD4	GS10_1	ISO-25	QH09_201	SC10_18	XZ09_15	XZ-ML_3	
CUHK: CD5	GS10_10	ISO-26	QH09_210	SC10_4	XZ09_32	XZ-ML_191	
CUHK: CD6	GYOKUJU	ISO-27	QH09_229	SC10_21	XZ09_46	XZ-NQ_74	
CUHK: CD7	GZUH2010LL4	ISO-28	QH09_255	SC-2	XZ09_48	XZ-NQ_80	

Table S2: Comparisons of the ribosomal 28S sequence of LKHE01000582 with the transcriptome sequences of natural *C. sinensis*.

Query segment of LKHE01000582	Subject transcriptome sequence segment		% Similarity	Note
	Accession #	Range & direction		
2,623→3,269	GAGW01000468	1←647	99.7% (645/647)	Overlapped
	GAGW01013978	1→248	98.0% (244/249)	
3,410→4,074	GAGW01000467	4←608	100% (605/605)	Overlapped with the longest transcript GAGW01000467
	GAGW01014526	197←467	100% (271/271)	
	GAGW01013228	263←517	100% (255/255)	
	GAGW01014061	279←486	100% (208/208)	
	GAGW01005959	1→202	100% (164/164)	
	GAGW01013560	1←330	99.8% (328/330)	
	GAGW01014446	186←472	99.7% (286/287)	
	GAGW01000457	1←242	99.6% (241/242)	
	GAGW01014495	215→470	99.6% (255/256)	
	GAGW01016267	1←284	99.6% (284/285)	
	GAGW01013436	304←507	99.5% (203/204)	
	GAGW01014446	1→192	99.5% (191/192)	
	GAGW01010050	1←192	99.5% (191/192)	
4,062→5,001	GAGW01009114	819←1,061	100% (243/243)	Overlapped or partially overlapped
	GAGW01000464	1←213	100% (213/213)	
	GAGW01004820	880→983	100% (104/104)	
	GAGW01000465	1→720	99.9% (719/720)	
	GAGW01016599	1←215	99.1% (214/216)	
	GAGW01014489	1→458	98.0% (451/460)	
	GAGW01006728	1,669→1,792	96.8% (122/126)	
	GAGW01007462	1→101	100% (101/101)	
5,341→6,454	GAGW01000465	720→1,682	99.5% (961/966)	Overlapped with the longest transcript GAGW01000465
	GAGW01007932	2←156	99.3% (151/152)	

Table S3: Comparison of the genome segment sequences of LKHE01000676 with other genome sequences.

Query segment of LKHE01000676	Subject genome sequence segment		% Similarity
	Accession #	Range & direction	
827→1,351	JAAVMX010000004	483→909	91.9% (398/433)
	JAAVMX010000001	4,008,144←4,008,663	86.6% (459/530)
	ANOV01006005	1,600←2,119	86.6% (459/530)
	LWBQ01000135	140,304→140,677	82.0% (319/389)
6,664→7,561	JAAVMX010000004	6,001→6,898	100% (898/898)
	LWBQ01000084	217,277←218,172	96.1% (864/899)
	ANOV01000289	16,637←17,532	96.0% (863/899)
19,006→20,474	JAAVMX010000004	19,343→19,811	100% (1,469/1,469)
	LWBQ01000084	205,382←206,468	86.9% (971/1,117)
	ANOV01000289	5,139←6,571	85.9% (1,280/1,489)
72,454→73,040	JAAVMX010000004	71,814→72,400	100% (587/587)
	LWBQ01000084	123,172←123,768	95.0% (570/600)
	ANOV01001676	16,560←17,156	
75,381→75,736	JAAVMX010000004	74,741→75,096	100% (356/356)
	LWBQ01000084	120,501←120,849	95.3% (343/360)
	ANOV01001676	13,889←14,237	94.7% (341/360)
86,199→86,377	JAAVMX010000004	85,559→85,737	100% (179/179)
	LWBQ01000084	109,827←110,001	91.7% (243/260)
	ANOV01001676	3,226←3,400	
88,424→89,005	JAAVMX010000004	87,784→88,365	100% (582/582)
	LWBQ01000084	107,219←107,782	91.2% (541/593)
	ANOV01001676	615←1,183	93.7% (554/591)

Table S4: Comparisons of the genome segment sequences of LWBQ01000028 with transcriptome sequences.

Query segment of LWBQ01000028	Subject transcriptome sequence segment		% Similarity	Note
	Accession #	Range & direction		
54,439→54,768	GAGW01006422	408←737	97.9% (323/330)	
	GCQL01004018	984←1,296	90.9% (300/330)	
73,018→74,296	GAGW01010266	6→445	87.9% (427/486)	
	GCQL01005624	2,569→3,842	92.0% (1,214/1,320)	
156,386→157,384	GAGW01006845	1,413→2,320	94.8% (906/956)	} Overlapped
	GCQL01001211	4,338→5,277	93.0% (929/999)	
	GCQL01019059	901→1,833	92.6% (926/1,000)	
158,450→158,863	GAGW01007074	1,028→1,390	87.7% (363/414)	} Overlapped
	GCQL01019059	2,899→3,261	87.4% (362/414)	
	GCQL01001211	6,343→6,711	82.1% (345/420)	
199,589→202,182	GAGW01008721	4→1,174	98.5% (1,154/1,172)	} Overlapped
	GAGW01007466	819←1,731	93.5% (906/969)	
	GCQL01006423	2,277←4,758	95.3% (2,473/2,594)	
	GCQL01019079	1,651←4,122	93.9% (2,442/2,601)	
205,025→205,882	GAGW01006682	1→814	94.5% (811/858)	} Overlapped
	GCQL01014655	2,604←3,417	94.6% (812/858)	
	GCQL01018712	211→1,008	93.0% (798/858)	
229,628→230,766	GAGW01006686	1,014←1,497	90.6% (483/533)	
		1,499←1,986	90.2% (488/541)	
	GCQL01008384	2,225←3,262	91.0% (1,037/1,140)	
249,156→250,435	GAGW01001103	862←2,056	93.4% (1,195/1,280)	
	GCQL01002205	1,847←3,041	93.3% (1,194/1,280)	
272,074→273,027	GAGW01006656	1,778→2,479	98.2% (698/711)	
		2,354→2,611	95.5% (255/267)	

	GCQL01008726	3,578←4,417	86.8% (828/954)	
339,248→339,679	GAGW01006419	775←1,092	93.8% (405/432)	
	GCQL01012312	1,667→2,084	93.8% (405/432)	
365,433→366,630	GAGW01011770	1→611	99.8% (610/611)	} Overlapped
	GAGW01014694	1→460	99.6% (458/460)	
	GCQL01014229	71←1,158	89.8% (1,078/1,200)	} Overlapped
	GCQL01017579	6,952→8,039	89.8% (1,078/1,200)	
416,688→417,407	GCQL01000179	1,861→2,473	84.6% (609/720)	
501,303→502,868	GAGW01007620	25→1,630	95.5% (1,541/1,613)	
	GCQL01010520	28←1,261	96.2% (1,190/1,237)	

Table S5: Comparison of the genome segment sequences of LWBQ01000037 with other genome sequences.

Query segment of LWBQ01000037	Subject genome sequence segment		% Similarity
	Accession #	Range & direction	
3,000→4,250	LKHE01003354	9,029→10,303	82.4% (1,079/1,309)
	JAAVMX010000003	11,246,204→11,247,478	82.4% (1,079/1,309)
	ANOV01025040	16→657	82.0% (625/703)
21,357→21,691	LKHE01003354	27,190→27,524	96.2% (325/338)
	JAAVMX010000003	11,262,877→11,263,211	96.2% (325/338)
	ANOV01008344	105←446	90.7% (313/345)
22,218→22,517	LKHE01001222	12,734←13,033	95.0% (285/300)
	JAAVMX010000006	5,963,852←5,964,151	95.0% (285/300)
	ANOV01021304	35→334	93.0% (279/300)
32,993→33,701	LKHE01000618	21,012→21,720	95.9% (680/709)
	JAAVMX010000004	8,620,917←8,621,614	95.4% (666/698)
	ANOV01006300	5,621←6,318	95.1% (664/698)
49,348→49,987	JAAVMX010000003	11,290,355→11,290,930	89.4% (572/640)
	LKHE01003354	55,164→55,860	87.6% (614/701)
	ANOV01021052	1,345←1,659	95.1% (291/306)
51,880→52,623	LKHE01003354	57,753→58,500	96.3% (724/752)
	JAAVMX010000003	11,292,823→11,293,570	96.3% (724/752)
	ANOV01021586	976→1,720	90.0% (677/752)
150,583→150,962	LKHE01003221	167,818←168,197	95.8% (364/380)
	JAAVMX010000001	4,177,123→4,177,502	95.8% (364/380)
	ANOV01005264	4,889←5,268	95.8% (364/380)
153,004→153,680	LKHE01003221	165,139←165,776	94.2% (638/677)
	JAAVMX010000001	4,179,544←4,180,181	94.2% (638/677)
	ANOV01005264	2,208→2,847	93.5% (635/679)
184,000→184,638	LKHE01003221	134,469←135,067	93.7% (599/639)
	JAAVMX010000001	4,210,254→4,210,852	93.7% (599/639)
	ANOV01001961	5,452←6,050	93.7% (599/639)
303,615→304,152	JAAVMX010000001	12,374,117→12,374,684	94.4% (508/538)
	LKHE01001481	4,604←5,096	91.6% (493/538)
	ANOV01014597	1,424→1,968	85.1% (464/545)
305,113→305,593	LKHE010034951	10,889←11,370	95.9% (462/482)
	JAAVMX010000002	5,251,254→5,251,734	95.4% (459/481)
	ANOV01003543	2111←2591	94.2% (454/482)
311,840→312,314	LKHE01003485	18,108→18,582	96.4% (458/475)
	JAAVMX010000002	10,070,414←10,070,889	96.2% (458/476)
	ANOV01016710	2,837←3,311	94.1% (447/475)
332,925→333,585	LKHE01002661	2,935←3,595	95.5% (631/661)
	JAAVMX010000006	4,798,403→4,799,063	95.5% (631/661)
	ANOV01003086	9,354→10,014	94.0% (621/661)
422,031→422,570	LKHE01002503	54,676←55,185	94.4% (510/540)
	JAAVMX010000001	21,035,423→21,035,932	94.4% (510/540)
	ANOV01009409	1,585→2,094	94.4% (510/540)

512,408→512,999	LKHE01003449	104,913←105,504	94.3% (558/592)
	JAAVMX010000004	4,373,490←4,374,081	94.3% (558/592)
	ANOV01018207	411→1,002	93.4% (553/592)

Table S6: Comparison of the genome segment sequences of LWBQ01000037 with transcriptome sequences.

Query segment of LWBQ01000037	Subject transcriptome sequence segment		% Similarity	Note
	Accession #	Range & direction		
6,232→7,009	GAGW01002594	44→768	80.0% (614/768)	
	GCQL01010165	1←207	88.3% (188/213)	
14,715→15,238	GAGW01008312	791→1,136	86.6% (465/537)	Overlapped
	GAGW01007593	206←836	81.9% (555/678)	
39,296→39,287	GAGW01016014	1→329	92.2% (309/335)	
	GCQL01008683	1→125	92.8% (116/125)	
48,955→49,073	GAGW01004410	314←1,094	85.2% (665/781)	
	GCQL01010547	19→117	85.3% (666/781)	
153,404→153,580	GAGW01007916	1,230←1,314	91.8% (78/85)	
	GCQL01009492	19→117	90.1% (91/101)	
238,826→239,489	GAGW01004410	417←1,080	86.3% (574/665)	
	GCQL01010547	2,115→2,781	86.4% (577/668)	
352,166→352,524	GAGW01005792	28→381	92.7% (332/358)	}Overlapped
	GCQL01017981	1→187	100% (187/187)	
	GCQL01005417	119←289	88.2% (165/187)	
389,632→389,985	GAGW01010267	520→666	100% (147/147)	
	GCQL01011582	1→347	86.3% (309/358)	
394,739→395,647	GAGW01006364	1,562←1,983	95.3% (403/423)	
	GCQL01008062	10←431	95.7% (405/423)	
	GCQL01011582	4,878→5,203	97.6% (319/327)	
410,957→411,111	GAGW01007528	352→502	87.9% (138/157)	
	GCQL01008907	611→761		
452,296→452,665	GAGW01011791	4←382	88.7% (338/381)	Overlapped
	GAGW01007067	910→1,288	84.8% (323/381)	

Table S7: Comparisons of the OSRC27 gene sequence JQ325719 (the Query) of strain NP10-2 with other OSRC27 gene sequences.

Accession #	<i>H. sinensis</i> strain	% Similarity vs. JQ325719 (strain NP10-2)
JQ227433	QH09-93	100% (574/574)
JQ227434	XZ07-176	100% (574/574)
JQ277435	SC09-37	100% (574/574)
JQ325605	GS09-121	100% (574/574)
JQ325618	QH07-197	100% (574/574)
JQ325640	SC09-36	100% (574/574)
JQ325641	SC09-47	100% (574/574)
JQ325654	SC09-190	100% (574/574)
JQ325671	XZ07-133	100% (574/574)
JQ325683	XZ08-59	100% (574/574)
JQ325687	XZ09-15	100% (574/574)
JQ325690	XZ09-48	100% (574/574)
JQ325696	XZ09-106	100% (574/574)
JQ325704	YN09-3	100% (574/574)
JQ325708	YN09-61	100% (574/574)
JQ325710	YN09-72	100% (574/574)
JQ325711	YN09-81	100% (574/574)
JQ325712	YN09-85	100% (574/574)
JQ325713	YN09-89	100% (574/574)
JQ325714	YN09-96	100% (574/574)
JQ325715	YN09-101	100% (574/574)

JQ325716	YN09-140	100% (574/574)
JQ325718	NP10-1	100% (574/574)
JQ325609	SC09-225	99.8% (573/574)
JQ325661	XZ05-6	99.8% (573/574)
JQ325693	XZ09-80	99.8% (573/574)
JQ325694	XZ09-95	99.8% (573/574)
JQ325613	GS09-311	99.7% (572/574)
JQ325634	QH09-210	99.7% (572/574)
JQ325717	ID10-1	99.7% (572/574)
JQ325614	GS09-327	99.5% (571/574)
JQ325675	XZ07-H1	97.2% (558/574)
JQ277436	XZ07-H2	97.0% (557/574)
JQ325707	YN09-51	94.6% (558/590)
JQ325709	YN09-64	94.6% (558/590)
JQ325706	YN09-22	94.0% (562/598)
JQ325705	YN09-6	93.8% (561/598)

Table S8: Comparisons of the genome segment sequences of ANOV01001374 and LWBQ01000131.

Query segment of ANOV01001374	Subject segment of LWBQ01000131	% Similarity
16,281→16,588	207,045←207,352	96.4% (297/308)
14,391→16,172	207,574←209,400	96.3% (1,760/1,828)
25,385→26,485	195,815←196,956	96.2% (1,099/1,142)
3,896→5,876	218,866←220,857	95.7% (1,933/2,020)
16,387→20,407	202,862←207,046	95.2% (3,992/4,193)
9,465→11,566	212,573←214,734	95.1% (2,072/2,179)
20,267→20,740	202,403→202,861	94.7% (449/474)
5,791→6,162	218,494←218,865	94.3% (352/373)
7,429→7,876	216,629←217,060	94.0% (421/448)
831→2,238	223,035←224,525	93.3% (1,391/1,491)
27,696→28,962	192,854←194,203	93.2% (1,266/1,359)
23,991→25,578	196,963←198,569	92.2% (1,527/1,656)
26,285→27,696	194,285←195,815	91.6% (1,403/1,531)
22,143→22,511	200,264→200,613	91.1% (336/369)
25,138→25,578	196,963←197,363	90.9% (401/441)
29,820→30,421	191,322←191,995	89.0% (600/674)

Table S9: Comparisons of the genome segment sequences of LKHE01002043 with other genome sequences.

Query segment of LKHE01002043	Subject genome sequence segment		% Similarity
	Accession #	Range & direction	
8,568→10,001	LWBQ01000085	291,697→293,111	93.4% (1,330/1,424)
	JAAVMX010000003	15,105,669→15,107,102	95.3% (1,367/1,434)
	ANOV01000934	1,557←2,975	93.1% (1,327/1,425)
9,956→11,323	LWBQ01000048	288,413→289,776	87.8% (1,201/1,368)
	JAAVMX010000003	15,107,057→15,108,424	94.7% (1,295/1,368)
	ANOV01007855	4,115←5,482	94.4% (1,292/1,368)
13,943→14,226	LWBQ01000048	286,984→287,268	88.8% (253/285)
	JAAVMX010000003	15,935,494→15,935,764	96.3% (261/271)
	ANOV01007855	1,215←1,498	96.1% (271/282)
15,054→16,251	LWBQ01000048	338,981→340,138	96.4% (1,155/1,198)
	JAAVMX010000003	15,112,157→15,113,324	97.4% (1,167/1,198)
	ANOV01022780	1→539	99.8% (538/539)
18,225→26,062	LWBQ01000048	342,326→350,181	95.1% (7,481/7,863)
	JAAVMX010000003	15,115,398→15,123,123	96.0% (7,537/7,848)
	ANOV01003645	2,738←8,407	87.9% (4,990/5,679)
19,622→19,903	LWBQ01000048	343,723→344,004	94.0% (265/282)
	JAAVMX010000003	15,116,695→15,116,976	96.8% (273/282)

	ANOV01005038	3,050←3,331	94.3% (266/282)
23,848→25,263	LWBQ01000048	347,964→349,380	94.9% (1,345/1,417)
	JAAVMX010000003	15,120,906→15,122,322	95.4% (1,354/1,417)
	ANOV01005037	9,723←11,144	94.2% (1,340/1,422)
51,944→52,614	LWBQ01000048	389,722→390,413	94.9% (657/692)
	JAAVMX010000003	15,161,860→15,162,551	
	ANOV01006908	5,574→6,265	
52,610→52,924	LWBQ01000035	412,590→412,903	97.5% (306/314)
	JAAVMX010000006	6,610,638←6,610,954	98.4% (312/317)
	ANOV01005521	807→1140	92.8% (310/334)
57,759→58,008	LWBQ01000048	408,815→409,057	96.4% (241/250)
	JAAVMX010000003	15,180,150→15,180,392	
	ANOV01012844	1,595→1,837	

```

AB067721          218  AACTTTCAACAACGGATCTCTTGGTTCTGGCATC  GATG  AAGAACGCAGCGAAATGCGAT
JAAVMX010000019  19563  -----
JAAVMX010000019  32207  -----
JAAVMX010000018  13540  -----
JAAVMX010000018  26235  -----
JAAVMX010000018  38930  -----
JAAVMX010000018  51626  -----
JAAVMX010000017  9478   -----
JAAVMX010000017  22122  -----
JAAVMX010000017  34766  -----
JAAVMX010000017  47410  -----
JAAVMX010000008  1359   -----
JAAVMX010000008  13982  -----
JAAVMX010000002  18689076 -----
LWBQ01000008     991956 -----
LKHE01000582     2291   -----
ANOV01021709     1055   -----
JAAVMX010000002  18702254 ----- C-----
JAAVMX010000018  858    ----- A-----
JAAVMX010000019  6396   ----- CAC-----
JAAVMX010000019  44889  -----

AB067721          277  AAGTAATGTGAATTGCAGAATTCAGTGAACCA  TCGAATC  TTTGAACG  CACATTGC  GC
JAAVMX010000019  19622  -----
JAAVMX010000019  32266  -----
JAAVMX010000018  13599  -----
JAAVMX010000018  26294  -----
JAAVMX010000018  38989  -----
JAAVMX010000018  51685  -----
JAAVMX010000017  9417   -----
JAAVMX010000017  22063  -----
JAAVMX010000017  34707  -----
JAAVMX010000017  47351  -----
JAAVMX010000008  1417   -----
JAAVMX010000008  14041  -----
JAAVMX010000002  18689135 -----
LWBQ01000008     992015 -----
LKHE01000582     2350   -----
ANOV01021709     1114   -----
JAAVMX010000002  18702314 ----- T-----
JAAVMX010000018  918    ----- C----- C-----
JAAVMX010000019  6453   ----- C----- C-----
JAAVMX010000019  44948  -----

AB067721          334  CCGCCAGCACTCTGGCGGGCATGC  CTGTCCGAGCGTCATC
JAAVMX010000019  19679  -----
JAAVMX010000019  32323  -----
JAAVMX010000018  13656  -----
JAAVMX010000018  26353  -----
JAAVMX010000018  39048  -----
JAAVMX010000018  51742  -----
JAAVMX010000017  9362   -----
JAAVMX010000017  22006  -----
JAAVMX010000017  34650  -----
JAAVMX010000017  47294  -----
JAAVMX010000008  1475   -----
JAAVMX010000008  14101  -----
JAAVMX010000002  18689192 -----
LWBQ01000008     992075 -----
LKHE01000582     2407   -----
ANOV01021709     1171   -----
JAAVMX010000002  18702369 -----
JAAVMX010000018  973    -----
JAAVMX010000019  6511   -----
JAAVMX010000019  45005  ----- CT-----

```

Figure S1: Alignment of the 5.8S gene sequences of *H. sinensis* strains. AB067721 is 5.8S gene portion of the PCR-amplified ITS1-5.8S-ITS2 sequences of *H. sinensis* strain GYOKUJU [13]. LWBQ01000008, LKHE01000582, ANOV01021709 and JAAVMX010000002/JAAVMX010000008/JAAVMX010000017–JAAVMX010000019 are segments of the assembled genome sequences of *H. sinensis* strain ZJB12195, 1229, Co18, and IOZ07, respectively [26,28,35-36]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps. Four variable segment sequences of 5.8S genes with scattered mutant alleles are listed at the bottom lines of the figure.


```

JM973579          4  CCCGGGCGAGAGCCCGCCAGTCGGAGCGCCTGCAGCGGAACTCTGGCGGGCCGCC
LKHE01000582     5255 -----
JAAVMX010000019 22527 -----
ANOV01022831     954 -----

JM973579          64  ACCACGCGTGGCGCCGCGAGCTTGGTTGAGCTTGACTCTAGTTTACATTGTGAAAAG
LKHE01000582     5315 -----
JAAVMX010000019 22587 -----
ANOV01022831     1014 -----
GAGW01000465     720 -----

JM973579          124 ACATAGGAGGTGTAGAATAGGTGGGAGCCTCGGCGCCGGTGAATACCCTACTCTCTATT
LKHE01000582     5375 -----
JAAVMX010000019 22647 -----
ANOV01022831     1074 -----
GAGW01000465     754 -----

JM973579          184 GtttttttACTTATTCAATGAAGCGGGGCTGGATTTTCGT          CCAACTTCTGG
LKHE01000582     5435 --C-----
JAAVMX010000019 22707 --C-----
ANOV01022831     1110 AG-AGAGA-GG--A-----T-          -ATAT-----A---AAGCATTAC-A---
GAGW01000465     839 -----

JM973579          235 TCTTAAGGTCTCCGCGGGCCGACCCGGGTTGAAGACATTGTCAGGTGGGGAGTTGGCT
LKHE01000582     5486 -----
JAAVMX010000019 22758 -----
ANOV01022831     1165 -----
GAGW01000465     864 -----

JM973579          295 GGGGCGGCACATCTGTTAAACCATAACGCAGGTGTCCTAAGGGGGGCTCATGGAGAACAG
LKHE01000582     5546 -----
JAAVMX010000019 22818 -----
ANOV01022831     1225 -----
GAGW01000465     923 -----

JM973579          355 AAATCTCCAGTGGAAACAAAAGGGTAAAAGTCCCCTTGATTTTGATTTTCAGTGTGAATAC
LKHE01000582     5605 -----
JAAVMX010000019 22878 -----
ANOV01022831     1285 -----
GAGW01000465     982 -----

JM973579          415 AAACCATGAAAGTGTGGCCTATCGATCC
LKHE01000582     5666 -----
JAAVMX010000019 22938 -----
ANOV01022831     1345 -----
GAGW01000465     1043 -----

```

Figure S2: Alignment of the unassembled shotgun 28S gene JM973579 of *H. sinensis* strain YN07-8 with assembled genome and transcriptome sequences. JM973579 is a unassembled shotgun 28S gene sequence of strain YN07-8 [39]. LKHE01000582, JAAVMX010000019, and ANOV01022831 are segments of assembled shotgun 28S gene sequences of *H. sinensis* strains 1229, IOZ07, and Co18, respectively [26,28,36]. GAGW01000465 is a 28S gene segment of assembled transcriptome assemblies of natural *C. sinensis* [53]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.

```

ANOV01021101     181  CGAGGGGAGCCCTTCCGGACGGGGCTGCTCAACGAGCTCTACCCCGGCTTCAAGCGAGT
LKHE01000642     22559 -----
JAAVMX010000012 437727 -----

ANOV01021101     241  CGCCGCCATCCTCGCCGACGTCGGCTTC  CTCCAACACCCTGCCGTGGGGCCGGGGGAG
LKHE01000642     22619 -----
JAAVMX010000012 437667 -----

ANOV01021101     299  ATAGGAGAGGGCGAGGGACGAGTACGAGAGGATGCCGGGGGGCCCGTCGAGGCG  CGCA
LKHE01000642     22670 CA-----G-T--CC--- --CA-C-C --C- -----ACCT-GT--TA-CT-T--T--T-
JAAVMX010000012 437616 CA-----G-T--CC--- --CA-C-C --C- -----ACCT-GT--TA-CT-T--T--T-

ANOV01021101     358          GGAACTCG  TCGTCGGCACCTTTCACGCGTCGGACATTCTTTCATCTTCTA
LKHE01000642     22727 CCTGTAC--C--C-CCT-T-----
JAAVMX010000012 437559 CCTGTAC--C--C-CCT-T-----

ANOV01021101     409  CGGCATCCTGCCCAACAAGGCCATGCAGAGCTGCCGCACCTACTACTTCAACTTCCTCCA
LKHE01000642     22787 -----
JAAVMX010000012 437499 -----

ANOV01021101     469  CAACCTCGACCCCAACAAGGGCGTGGGCGGCTATGCAGAGTGGCCCGGTTGGCGGGACGG
LKHE01000642     22847 -----
JAAVMX010000012 437439 -----

ANOV01021101     529  GAACCGGCTCATGTGGTTCAAGTCGGCCTTTGAAAACGAAATCATCGACGACGACTTCCG
LKHE01000642     22907 -----
JAAVMX010000012 437379 -----

```

Figure S3: Alignment of the genome assembly segments ANOV01021101, LKHE01000642, and JAAVMX010000012. ANOV01021101, LKHE01000642, and JAAVMX010000012 are segments of the genome assembly sequences of *H. sinensis* strains Co18, 1229 and IOZ07, respectively [26,28,36]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.

```

LWBQ01000028      226274 TGCAAAAGTCTCGGAGAACCAAGAGGCAACTCTGGGATAGCCTGACCATCAGTGGTGCCT
LKHE01002847      40829 -----
ANOV01006525      3537 -----
JAAVMX010000008  1837159 -----

LWBQ01000028      226334 CGGTGCGTCG AGC GCACGCCCGCCCATCTCAGATGAAAGGCTGACCATTGGCAGNNNN
LKHE01002847      40889 --A -TC---A-C- -----
ANOV01006525      3477  --A -TC---A-C- -----
JAAVMX010000008  1837099 --AGT- --- -A-C-----

LWBQ01000028      226392 NNNNNNGCCGTTTGGCAGCCGTCACGAGGGCCTTTACTCTCGTCTACACTCTCGCCCTCC
LKHE01002847      40936 -----
ANOV01006525      3430 -----
JAAVMX010000008  1837052 -----

LWBQ01000028      235271 ACCGactactactactactgactacttgctactactactactCCCGCATACATAcccccc
LKHE01002847      49806 -----
JAAVMX010000008  1828183 -----

LWBQ01000028      235331 cccnnnnnnnnnnnnccccccccacacctcccccccccccTCCTCTCTTCTTCGCCTCTAA
LKHE01002847      49866 ---
JAAVMX010000008  1828123 ---

LWBQ01000028      235391 TAAGCCCGCCAAACAAATCAACCACCCTTACGGCGTCTCTAGCAGCCACCTTGCGCGCCC
LKHE01002847      49891 -----
JAAVMX010000008  1828098 -----

LWBQ01000028      290668 CCAA AAAGCCAAGCGAAAATCCTTACTCACGTCACGCAGGACCTGGGAGGCCAGACCAG
LKHE01002847      104998 ----G-----
ANOV01006525      6145 ----G-----
JAAVMX010000008  1773217 ----G-----

LWBQ01000028      290727 CATCCTGGCGGTGCTNNNNNNNNNNCTACTACGCCAACACGGCCGCCGTCATCTTCGTGG
LKHE01002847      105058 ----- -C-CC-A-- GGCGGTG -----
ANOV01006525      6085 ----- -C-CC-A-- TGGCGGTG -----
JAAVMX010000008  1773157 ----- -C-CC-A-- GGCGGTG -----

LWBQ01000028      290787 TCGACTCGACCGACATTGAACGACTGCAGACGGCAGCCGACGAGCTGGCCGCCATGCTGA
LKHE01002847      105114 -----
ANOV01006525      6029 -----
JAAVMX010000008  1773101 -----

```

Figure S4: Alignment of the genome assembly segments LWBQ01000028, ANOV01000226/ANOV01006525, LKHE01002847, and JAAVMX010000008. LWBQ01000028, ANOV01000226/ANOV01006525, LKHE01002847, and JAAVMX010000008 are segments of the genome assembly sequences of *H. sinensis* strains ZJB12195, Co18, 1229, and IOZ07, respectively [26,28,35-36]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.

```

LWBQ01000037      150583 tttttttGTTTTATATTGTCGGATTGTCGGTCTTTTAGTAAacccccccTCCTTACGGAG
LKHE01003221      168197 -----TC-T-TTCCACG-----
ANOV01005264       5268 -----TC-T-TTCCACG-----
JAAVMX010000001  4177123 -----TC-T-TTCCACG-----

LWBQ01000037      150643 CAAGGCGAAGAAGGAGGGGGATGCCGGATTTAAGGATTTAATAGTAGGGTCTGTATTCTA
LKHE01003221      168137 -----CCCCCT-----
ANOV01005264       5208 -----CCCCCT-----
JAAVMX010000001  4177183 -----CCCCCT-----

LWBQ01000037      168381 CCTCTAAGAGCGCAGGTGGTAAGGCGGTTAAGTAGGGCAGAAAAAGCCTACGAGATACG
GAGW01007916      1367 ----G-GGT-T-----C-G-----C---G-----C---T-----
GCQL01009492       1 -----C---G-----TA-----

LWBQ01000037      168441 TGTTTTTCGTAGGG CGCGGCAGCGCTTTTCTAGATCCACGATTACTTCCTTTTTTAGGC
GAGW01007916      1307 -----A---A---GC-----C-----
GCQL01009492       39 -----T-A-A---T-A-GA---T---T-----T-----

LWBQ01000037      168500 CCTCTTTAAGCCTTGTTTTCTATGCAACGTCGCACTATATTGTGACGTGCTGTTTTTTA
GAGW01007916      1248 -----G-----G-----C-C--C---
GCQL01009492       99 -----G-----A--- --G-----

LWBQ01000037      422191 TACCTGAAGCTGCTCGCCAGCTCCTCCAGACTCAACCATGGCTGCCGCGACTCAAAGTC
LKHE01002503       55025 -----
ANOV01009409       1934 -----
JAAVMX010000001  21035583 -----
GAGW01009411       790 -----
GCQL01006072       456 -----

LWBQ01000037      422251 TCCACCATGGCTGCCCATGGCTGCCGTCGTCACGTTTTGCGGGGCCCTGCCGCGCACGC
LKHE01002503       54985 -----
ANOV01009409       1894 -----
JAAVMX010000001  21035623 -----
GAGW01009411       830 -----
GCQL01006072       493 -----

LWBQ01000037      424561 TTGTGCGCGTGCATGTACCTCGAGACAGTGTGTTTTTCCACGCGGTACTAGGGGGACT
LKHE01002503       52595 -----
ANOV01009408       2371 -----
JAAVMX010000001  21038058 -----
GAGW01009978       696 -----
GCQL01006297       1211 -----

LWBQ01000037      424621 AAGAATTCGTCCCTCAACCCCTCCTTCTATTAG CCG
LKHE01002503       52535 -----GG--GGGGGGGACA---ACCAACAAAAAGC
ANOV01009408       2311 -----GG--GGGGGGGACA---ACCAACAAAAAGC
JAAVMX010000001  21038118 -----GG--GGGGGGGACA---ACCAACAAAAAGC
GAGW01009978       696 --A-----
GCQL01006297       1227 --A-----

LWBQ01000037      424658 CTTAGCCGCACCACCTATTTTCTAGTAAAGCAA
LKHE01002503       52475 CCAGCGACTCGCAAAAAACAGGTA-----
ANOV01009408       2251 CCAGCGACTCGCAAAAAACAGGTA-----
JAAVMX010000001  21038178 CCAGCGACTCGCAAAAAACAGGTA-----
GAGW01009978       794 -----T-----
GCQL01003755       132 -----A---T-G-----

LWBQ01000037      424693 GTATCCCGATTTTTTCGGGTAGATGC GTAGCTACCTT
LKHE01002503       52415 -----ACATTTTC-----
ANOV01009408       2191 -----ACATTTTC-----
JAAVMX010000001  21038238 -----ACATTTTC-----
GAGW01009978       829 ---C-----
GCQL01003755       97 ---C-----

```

Figure S5: Alignment of the genome assembly segment LWBQ01000037 with other genome and transcriptome sequences. LWBQ01000037, ANOV01001961/ANOV01005264/ANOV01009409/ANOV01013329, LKHE01002503/ LKHE01003221, and JAAVMX010000001/JAAVMX010000002 are segments of the assembled genome sequences of *H. sinensis* strains ZJB12195, Co18, 1229, and IOZ07, respectively [26,28,35-36]. GCQL01006072/GCQL01009492 and GAGW01007916/GAGW01009411/GAGW01009978 are segments of the transcriptome assemblies of strain L0106 and natural *C. sinensis* [33,53]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.

```

YN09-140 JQ325484 61 CGCATGGTGCCTCGCGAGGCAATTCCTCTCGCCCATGATGTGATTACTGAGGATGCCATC
XZ09-80 JQ325461 61 -----
YN09-3 JQ325472 61 -----
QH09-164 JQ325398 61 -----
QH09-151 JQ325397 61 -----
YN09-51 JQ325475 61 -----
SC09-37 JQ277391 61 -----
YN09-89 JQ325481 61 -----
SC09-36 JQ325408 61 -----
XZ06-124 JQ277392 61 -----
XZ07-H2 JQ325443 61 -----
XZ12-16 KM197544 61 -----
XZ05-8 JQ325431 61 -----
XZ07-H1 JQ325442 61 -----

YN09-140 JQ325484 121 TGCAGGCACTGGCAAGAGCCGTTTACGCTCGCAAGCCGATGCTCATTTTAGACGACGCTC
XZ09-80 JQ325461 121 -----
YN09-3 JQ325472 121 -----
QH09-164 JQ325398 121 -----
QH09-151 JQ325397 121 -----
YN09-51 JQ325475 121 -----
SC09-37 JQ277391 121 -----
YN09-89 JQ325481 121 -----
SC09-36 JQ325408 121 -----
XZ06-124 JQ277392 121 -----
XZ07-H2 JQ325443 121 -----
XZ12-16 KM197544 121 -----
XZ05-8 JQ325431 121 -----
XZ07-H1 JQ325442 121 -----

YN09-140 JQ325484 181 TTAGCGCCCTTGACGGGCAACAGCGCAAGCCGCTCTCCTTCAGCTGCTTGGCCGTGATG
XZ09-80 JQ325461 181 -----
YN09-3 JQ325472 181 -----
QH09-164 JQ325398 181 -----
QH09-151 JQ325397 181 -----
YN09-51 JQ325475 181 -----
SC09-37 JQ277391 181 -----
YN09-89 JQ325481 181 -----
SC09-36 JQ325408 181 -----
XZ06-124 JQ277392 181 -----
XZ07-H2 JQ325443 181 -----
XZ12-16 KM197544 181 -----
XZ05-8 JQ325431 181 -----
XZ07-H1 JQ325442 181 -----

YN09-140 JQ325484 241 GGCTATTGCGAAAACCTGGGACCACAGTCGTCATGACTACCCACTCAGGTGAGCAGAAAC
XZ09-80 JQ325461 241 -----
YN09-3 JQ325472 241 -----
QH09-164 JQ325398 241 -----
QH09-151 JQ325397 241 -----
YN09-51 JQ325475 241 -----
SC09-37 JQ277391 241 -----
YN09-89 JQ325481 241 -----
SC09-36 JQ325408 241 -----
XZ06-124 JQ277392 241 -----
XZ07-H2 JQ325443 241 -----
XZ12-16 KM197544 241 -----
XZ05-8 JQ325431 241 -----
XZ07-H1 JQ325442 241 -----

YN09-140 JQ325484 301 CCACTCACTGAG GGAGGCTGCTACGTGTCGAGCAA GAATTCCTCGCT
XZ09-80 JQ325461 301 -----
YN09-3 JQ325472 301 -----
QH09-164 JQ325398 301 -----
QH09-151 JQ325397 301 -----
YN09-51 JQ325475 301 -----
SC09-37 JQ277391 301 -----
YN09-89 JQ325481 301 -----
SC09-36 JQ325408 281 -----
XZ06-124 JQ277392 301 TAGAAGACAT GGT -A-T
XZ07-H2 JQ325443 301 TAGAAGACAT GGT -A-T
XZ12-16 KM197544 301 TAGAAGACAT GGT -A-T
XZ05-8 JQ325431 301 TAGAAGACAT GGT -A-T
XZ07-H1 JQ325442 301 TAGAAGACAT GGT -A-T

YN09-140 JQ325484 348 GACTTGGTGGATGCCGAGTTGAGCATCTCGCCCTAGCCGACAGATTCTCGTACTAGAT
XZ09-80 JQ325461 348 -----
YN09-3 JQ325472 348 -----
QH09-164 JQ325398 348 -----
QH09-151 JQ325397 348 -----
YN09-51 JQ325475 348 -----
SC09-37 JQ277391 348 -----
YN09-89 JQ325481 348 -----
SC09-36 JQ325408 328 -----
XZ06-124 JQ277392 361 -----
XZ07-H2 JQ325443 361 -----
XZ12-16 KM197544 361 -----
XZ05-8 JQ325431 361 -----
XZ07-H1 JQ325442 361 -----

YN09-140 JQ325484 408 GGCACCGGAAGTGTTCGAGAACAGAGACAGTCGATGATCCGAAAGACAGAGCTTCGTC
XZ09-80 JQ325461 408 -----
YN09-3 JQ325472 408 -----
QH09-164 JQ325398 408 -----
QH09-151 JQ325397 408 -----
YN09-51 JQ325475 408 -----
SC09-37 JQ277391 408 -----
YN09-89 JQ325481 408 -----
SC09-36 JQ325408 388 -----
XZ06-124 JQ277392 421 -----
XZ07-H2 JQ325443 421 -----
XZ12-16 KM197544 421 -----
XZ05-8 JQ325431 421 -----
XZ07-H1 JQ325442 421 -----

YN09-140 JQ325484 468 CATGTGCGAAAACCGGGACACGCGGAACTGAAAACGCTCAAACTGTGTCCAGG
XZ09-80 JQ325461 468 -----
YN09-3 JQ325472 468 -----
QH09-164 JQ325398 468 -----
QH09-151 JQ325397 468 -----
YN09-51 JQ325475 468 -----
SC09-37 JQ277391 468 -----
YN09-89 JQ325481 468 -----
SC09-36 JQ325408 448 -----
XZ06-124 JQ277392 481 -----
XZ07-H2 JQ325443 481 -----
XZ12-16 KM197544 481 -----
XZ05-8 JQ325431 481 -----
XZ07-H1 JQ325442 481 -----

```

Figure S6: Alignment of the OSRC14 marker gene sequences of *H. sinensis* strains. GenBank accession numbers for OSRC14 marker gene sequences (second column from left) of *H. sinensis* strains (far left column) are listed in this figure and provided in TABLE 6 [40-42]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.


```

NP10-2 JQ325719 61 TGACCCCAAGGCGCTCCAAAGCGAGCCGTACCGGAGACTCACCGAGCCCGCCATAGGGAA
YN09-140 JQ325716 61 -----
QH09-93 JQ277433 61 -----
ID10-1 JQ325717 61 -----
XZ07-H2 JQ277436 62 -----
YN09-22 JQ325706 61 -----T-----
YN09-6 JQ325705 61 -----T-----
YN09-64 JQ325709 61 -----T-----
YN09-51 JQ325707 61 -----T-----

NP10-2 JQ325719 121 CATCCCTATATTCACGTGAGTGACCCCTGGACGGGCACTCCAGCCCTGAGCCTGCGTG
YN09-140 JQ325716 121 -----
QH09-93 JQ277433 121 -----
ID10-1 JQ325717 121 -----
XZ07-H2 JQ277436 122 -----
YN09-22 JQ325706 121 -----G-----
YN09-6 JQ325705 121 -----G-----
YN09-64 JQ325709 121 -----G-----
YN09-51 JQ325707 121 -----G-----

NP10-2 JQ325719 181 GCCGGCACAGAGCCTGGCACCGCAGCAGG ACCCCCG
YN09-140 JQ325716 181 -----
QH09-93 JQ277433 181 -----
ID10-1 JQ325717 181 -----
XZ07-H2 JQ277436 182 -----T-----
YN09-22 JQ325706 181 -----CGCAGCACCGCAGCACCGCAGCGCC-----
YN09-6 JQ325705 181 -----CGCAGCACCGCAGCACCGCAGCGCC-----
YN09-64 JQ325709 181 -----G-----CGCAGCACCGCAGCACC-----
YN09-51 JQ325707 181 -----G-----CGCAGCACCGCAGCACC-----

NP10-2 JQ325719 217 AGACTAATGTCGGTTTTTCGAGTGCAGAGACACCGGCCCAAGGGCGGAAGCTTCGGCTGCG
YN09-140 JQ325716 217 -----
QH09-93 JQ277433 217 -----
ID10-1 JQ325717 217 -----A-----
XZ07-H2 JQ277436 215 -----C-----AT-----G-----C-----A-----
YN09-22 JQ325706 241 -----C-----
YN09-6 JQ325705 241 -----C-----
YN09-64 JQ325709 233 -----C-----C-----A-----
YN09-51 JQ325707 233 -----C-----C-----A-----

NP10-2 JQ325719 277 GAGAGAGGAACACGTCGTGCCACCACACAGAAACCTTCATCCGCTCTGCTTCCGGTAGC
YN09-140 JQ325716 277 -----
QH09-93 JQ277433 277 -----
ID10-1 JQ325717 277 -----G-----
XZ07-H2 JQ277436 275 -----G-----G-----
YN09-22 JQ325706 301 -----T-----C-----G-----
YN09-6 JQ325705 301 -----T-----C-----G-----
YN09-64 JQ325709 293 -----T-----GG-----C-----G-----
YN09-51 JQ325707 293 -----T-----GG-----C-----G-----

NP10-2 JQ325719 337 CTGCCGCCAAAGGAGAATAAAGAACCAAGGTCTCGCCGCGACTATGTGACCCCTCGTCTG
YN09-140 JQ325716 337 -----
QH09-93 JQ277433 337 -----
ID10-1 JQ325717 337 -----
XZ07-H2 JQ277436 335 -----T-----
YN09-22 JQ325706 361 -----C-----
YN09-6 JQ325705 361 -----C-----
YN09-64 JQ325709 353 -----C-----
YN09-51 JQ325707 353 -----C-----

NP10-2 JQ325719 397 CAGGAGCTGAATATCAACCGCTGGACAAACGTCACATAGACAACACTACATGGCTGCCAAG
YN09-140 JQ325716 397 -----
QH09-93 JQ277433 397 -----
ID10-1 JQ325717 397 -----
XZ07-H2 JQ277436 395 -----A-----
YN09-22 JQ325706 421 -----A-----C-----
YN09-6 JQ325705 421 -----A-----C-----
YN09-64 JQ325709 413 -----A-----C-----
YN09-51 JQ325707 413 -----A-----C-----

NP10-2 JQ325719 457 TGCATTTCCTCTACCTGAAACCGACCATCCCTGCTGGGCCACTTTGACCCCGAGCTG
YN09-140 JQ325716 457 -----
QH09-93 JQ277433 457 -----
ID10-1 JQ325717 457 -----
XZ07-H2 JQ277436 455 -----A-----
YN09-22 JQ325706 481 -----
YN09-6 JQ325705 481 -----
YN09-64 JQ325709 473 -----
YN09-51 JQ325707 473 -----

NP10-2 JQ325719 517 TTCGTTTCACACCTGACTGCTGGCCAGGGTGAGTTTGTCTTCACTGCTAGTCAACG
YN09-140 JQ325716 517 -----
QH09-93 JQ277433 517 -----
ID10-1 JQ325717 517 -----
XZ07-H2 JQ277436 515 -----C-----
YN09-22 JQ325706 541 -----C-----
YN09-6 JQ325705 541 -----C-----
YN09-64 JQ325709 533 -----C-----
YN09-51 JQ325707 533 -----C-----

```

Figure S8: Alignment of the OSRC27 marker gene sequences of *H. sinensis* strains. GenBank accession numbers for the OSRC27 gene sequences (second column from left) of *H. sinensis* strains (far left column) are listed in this figure and provided in TABLE S8 [39,41-42]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.

```

JQ325719      1  CTTTGAGCACGTCGAGCTCGCCGCCAACATCCCATCGCCTATCCCGTCTTGACGCCCTT
JQ325705      1  -----G-----
GCQL01013271 2117 -----
GCQL01015312 154 -----

JQ325719      61  TGACCCCAAGGCGCTCCAAAGCGAGCCGTACCGGAGACTACCGAGCCCGCCATAGGGAA
JQ325705      61  -----T-----
GCQL01013271 2057 -----
GCQL01015312  94  -----

JQ325719      121 CATCCCTATATTCACGTGAGTGACCCCTGGACGCGGCACTCCAGCCCTGAGCCTGCGTG
JQ325705      121 -----G-----
GCQL01013271 1997 -----
GCQL01015312  34  -----

JQ325719      181 GCCGGCACAGAGCCTGGCACCGCAGCAGC          ACCCCCG
JQ325705      181 -----CGCAGCACCGCAGCACCGCAGCGCC-----

JQ325719      217 AGACTAATGTCGGTTTTTCGAGTGCAGAGACACCGGCCCAAGGCGGGAAGCTTCGGCTGCG
JQ325705      241 -----C-----
GCQL01013271 1985 -----C-----

JQ325719      277 GAGAGAGGAACACGTCGTGCCACCACACAGAAACCTTCATCCGCCTCTGCTTCGGGTAGC
JQ325705      301 -----T-----C-----G-----
GCQL01013271 1948 -----

JQ325719      337 CTGCCGCCAAAGGAGAATAAAGAACCAAAGGTCTCGCCGCGACTATGTGACCCCTCGTCTG
JQ325705      361 -----C-----
GCQL01013271 1888 -----

JQ325719      397 CAGGAGCTGAATATCAACCGCTGGACAAACGTCAACATAGACAACACTACATGGCTGCCAAG
JQ325705      421 -----A-----C-----
GCQL01013271 1828 -----

JQ325719      457 TGCATTTCTCTCTACCTGGAAACCGACCATCCCCTGCTGGGCCACTTTGACCCCGAGCTG
JQ325705      481 -----
GCQL01013271 1768 -----

JQ325719      517 TTCGTTTCACACCTGACTGCTGGCCAGGGTGAGTTTTGCTCTTCACTGCTAGTCAACG
JQ325705      541 -----C-----
GCQL01013271 1708 -----

```

Figure S9: Alignment of the OSRC27 marker gene sequences with transcriptome sequences GCQL01013271 of *H. sinensis* strain L0106. The OSRC14 marker gene sequences JQ325719 and JQ325706 of *H. sinensis* strains NP10-2 and YN09-22, respectively, aligned with the transcriptome assemblies GCQL01013271 and GCQL01015312 of *H. sinensis* strain L0106 [33,39,41-42,]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.

```

YN07-8      JM973601  513  CTGGCCCTTCTCCCTGACAGCAAGCTGCTGGCTTACAGCGCCGAATCGTGACGAGAGCCCG
SC09-37    JQ277447  301  -----
YN09-61    JQ325824  301  -----
SC09-47    JQ325757  298  -----G-----T-----
XZ07-11    JQ325784  301  -----G-----T-----
XZ05-12    JQ325780  298  -----G-----T-----
ID10-1     JQ325833  298  -----G-----T-----
SC09-77    JQ325760  298  -----G-----T-----
YN09-64    JQ325825  301  -----T-----G-----ATA-T-----
XZ08-4     JQ325792  301  -A-----T-----T-----
YN09-22    JQ325822  298  -----T-----G-----ATA-T-----
XZ07-176   JQ277445  301  -----C-----AT-----C-----
GS09-121   JQ325721  301  -A-----T-----T-----
YN09-3     JQ325820  301  -----C-----AT-----C-----
GS09-229   JQ325726  301  -----C-----AT-----C-----

YN07-8      JM973601  573  ATCATTCAACTCTGGGACCTGCAACAGGCTCCGCTCTTCATGCTTTGAGGGTCATGGC
SC09-37    JQ277447  361  -----
YN09-61    JQ325824  361  -----
SC09-47    JQ325757  358  -----C-----T-----
XZ07-11    JQ325784  361  -----T-----A-----C-----T-----
XZ05-12    JQ325780  358  -----C-----T-----
ID10-1     JQ325833  358  -----C-----T-----
SC09-77    JQ325760  358  -----C-----T-----
YN09-64    JQ325825  361  -----T-----C-----T-----
XZ08-4     JQ325792  361  -----C-----T-----
YN09-22    JQ325822  358  -----T-----C-----T-----
XZ07-176   JQ277445  361  -----C-----C-----T-----
GS09-121   JQ325721  361  -----C-----C-----T-----
YN09-3     JQ325820  361  -----C-----C-----T-----
GS09-229   JQ325726  361  -----C-----C-----T-----

YN07-8      JM973601  633  AAGCGCATCCAGGCCATCACATTCTCTCTGACAGCAAAACGCTGGCTTCAGTTCCAAT
SC09-37    JQ277447  421  -----
YN09-61    JQ325824  421  -----
SC09-47    JQ325757  418  -----T-----A-----T-----AC-----T-----
XZ07-11    JQ325784  421  -A-----AG-----C-----A-----A-----
XZ05-12    JQ325780  418  -----T-----A-----T-----A-----AC-----T-----
ID10-1     JQ325833  418  -----T-----A-----T-----AC-----T-----
SC09-77    JQ325760  418  -----T-----A-----T-----AC-----T-----
YN09-64    JQ325825  421  -A-----AG-----T-----AC-----T-----
XZ08-4     JQ325792  421  -----G-----A-----A-----
YN09-22    JQ325822  418  -----T-----A-----T-----AC-----T-----
XZ07-176   JQ277445  421  -A-----AG-----C-----A-----
GS09-121   JQ325721  421  -----G-----A-----A-----
YN09-3     JQ325820  421  -A-----AG-----C-----A-----
GS09-229   JQ325726  421  -A-----AG-----C-----A-----

YN07-8      JM973601  693  GATGAGACTGTCAGGCTCTGGGATACGGTGACGGGCCATCAGGTCAGTTCCTGAGCTG
SC09-37    JQ277447  481  -----
YN09-61    JQ325824  481  -----
SC09-47    JQ325757  478  -----C-----C-----G-----
XZ07-11    JQ325784  481  -----C-----C-----G-----G-----
XZ05-12    JQ325780  478  -----C-----C-----G-----A-----
ID10-1     JQ325833  478  -----C-----C-----G-----
SC09-77    JQ325760  478  -----C-----C-----G-----
YN09-64    JQ325825  481  -----C-----C-----G-----
XZ08-4     JQ325792  481  -----A-----C-----C-----C-----
YN09-22    JQ325822  478  -----A-----C-----G-----
XZ07-176   JQ277445  481  -----C-----A-----
GS09-121   JQ325721  481  -----C-----C-----
YN09-3     JQ325820  481  -----C-----A-----
GS09-229   JQ325726  481  -----C-----A-----

YN07-8      JM973601  753  AAAGGAAAGCTCGCCACCTGTGTGGACCCCGTCTGGCCACCCGTGTGGCGTTCTCGCCT
SC09-37    JQ277447  541  -----
YN09-61    JQ325824  541  -----
SC09-47    JQ325757  538  ---A-----T-----
XZ07-11    JQ325784  541  ---A-----T-T-----
XZ05-12    JQ325780  538  ---A-----T-----
ID10-1     JQ325833  538  ---A-----T-----
SC09-77    JQ325760  538  ---A-----T-----
YN09-64    JQ325825  541  ---A-----T-----
XZ08-4     JQ325792  541  ---A-----T-C---A-C-A---G-A-G-G---C-----
YN09-22    JQ325822  538  ---A-----A---A-----
XZ07-176   JQ277445  541  ---A-----T-C---A-C-A---G-A-G-G---C-----
GS09-121   JQ325721  541  ---A-----A---A-----
YN09-3     JQ325820  541  ---A-----A---A-----
GS09-229   JQ325726  541  ---A-----A---A-----G-----

YN07-8      JM973601  813  GATGGCAGTTTGCTAGCACCGTATACACTTTTTCGCCC
SC09-37    JQ277447  601  -----
YN09-61    JQ325824  601  -----
SC09-47    JQ325757  574  -----T-----
XZ07-11    JQ325784  577  -----C---CT---A-----
XZ05-12    JQ325780  574  -----T-C---T-----
ID10-1     JQ325833  574  -----T-C---T-----
SC09-77    JQ325760  574  -----T-C---T-----
YN09-64    JQ325825  577  -----C---CT---A-----
XZ08-4     JQ325792  601  -----T-----
YN09-22    JQ325822  574  -----T-C---GC-----
XZ07-176   JQ277445  589  -----T-----
GS09-121   JQ325721  601  -----T-----
YN09-3     JQ325820  589  -----T-----
GS09-229   JQ325726  589  -----T-----

```

Figure S10: Alignment of the unassembled genome sequence JM973601 with the OSRC32 marker gene sequences. JM973601 is a unassembled shotgun genome sequence of *H. sinensis* strain YN07-8 [39]. The strain information (the far left column) and accession numbers for the OSRC32 sequences (the second column from left) are listed in TABLE 6 [39]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.


```

KP090946      541 TCCGGCGTCGGCTCCAACGGCCGCGCCCATCACAAAGGCGACGCCCCGCTCTCCTCGCTC
GCQL01011182 957 -----
GAGW01003914 826 -----
LKHE01001673 5301 -----
ANOV01001719 21350 -----
JAAVMX010000006 5598556 -----
LWBQ01000045 386192 -----

KP090946      601 ACCAACTTCACCACGCACGACTACGAGAACCCAGCAGCATCTCCGCCTGCAGCCCAAG
GCQL01011182 1017 -----
GAGW01003914 766 -----
LKHE01001673 5361 -----
ANOV01001719 21290 -----
JAAVMX010000006 5598496 -----
LWBQ01000045 386239 -----
                    ---C-G-AT---

KP090946      661 CCGGCTACAAGCAGTGGGTCGCGCAGGCGGGCAGCTCATTGCCGATCTCCTCACCTGC
GCQL01011182 1077 -----
GAGW01003914 706 -----
LKHE01001673 5421 -----
ANOV01001719 21230 -----
JAAVMX010000006 5598436 -----
LWBQ01000045 386268 -----
                    T--C-

KP090946      721 GCCGGCGCCGACCACATCATCACCATGGACCTCCAGACCCCCAGTACCAGGGCTTCTTC
GCQL01011182 1137 -----
GAGW01003914 646 -----
LKHE01001673 5481 -----
ANOV01001719 21170 -----
JAAVMX010000006 5598376 -----
LWBQ01000045 386328 -----
                    T--A-

KP090946      781 GACGTGCCCGTCGACAACCTCTACGGGAAGCCCCTGCTGCAAAACTACATCGTGAAGCAC
GCQL01011182 1197 -----
GAGW01003914 586 -----
LKHE01001673 5541 -----
ANOV01001719 21110 -----
JAAVMX010000006 5598316 -----
LWBQ01000045 386388 -----

KP090946      841 ATTCCCAGCTACAAGGAGGCCGTCGTCGTCTCTCCGACGCTGGCGGGGCAAAGCGAGCC
GCQL01011182 1257 -----
GAGW01003914 526 -----
LKHE01001673 5601 -----
ANOV01001719 21050 -----
JAAVMX010000006 5598256 -----
LWBQ01000045 386448 -----
                    ---G---

KP090946      901 TCGGCCATTGCCGACGACCTCGACATGAAGTTTGCTCTCATCCACA
GCQL01011182 1317 -----
GAGW01003914 466 -----
LKHE01001673 5661 -----
ANOV01001719 20990 -----
JAAVMX010000006 5598196 -----
LWBQ01000045 386508 -----
                    AGGTAGTGCCGCGCT
                    AGGTAGTGCCGCGCT
                    AGGTAGTGCCGCGCT
                    AGGTAGTGCCGCGCT

KP090946      947 AGGAACGCCGTCATCAAGTTCAAC
GCQL01011182 1363 -----
GAGW01003914 454 -----
LKHE01001673 5721 -----
ANOV01001719 20930 -----
JAAVMX010000006 5598136 -----
LWBQ01000045 386568 -----
                    CCAGCGTTCCGTCGGGACTCCGCTCACTTTACTC-
                    CCAGCGTTCCGTCGGGACTCCGCTCACTTTACTC-
                    CCAGCGTTCCGTCGGGACTCCGCTCACTTTACTC-
                    CCAGCGTTCCGTCGGGACTCCGCTCACTTTACTC-

KP090946      973 GAGCAACGCAATGCCAGCATGATGCTCGTCGGCGACATTACCGACCGCTCTGCATCCTC
GCQL01011182 1389 -----
GAGW01003914 428 -----
LKHE01001673 5781 -----
ANOV01001719 20870 -----
JAAVMX010000006 5598076 -----
LWBQ01000045 386628 -----

KP090946      1033 GTCGACGATCTTGCCGATACCGCAAACACTGTTACTCGTGCCCAAGCT
GCQL01011182 1449 -----
GAGW01003914 368 -----
LKHE01001673 5841 -----
ANOV01001719 20810 -----
JAAVMX010000006 5598016 -----
LWBQ01000045 386688 -----
                    GCTCAAGCGC

KP090946      1083 GCTCAAGCGCTGGGCGCCACGCGCTCA
GCQL01011182 1499 -----
GAGW01003914 318 -----
LKHE01001673 5891 -----
ANOV01001719 20760 -----
JAAVMX010000006 5597966 -----
LWBQ01000045 386748 -----
                    GTGGGCGCCACGCGCTGTTACTCGTGCC-

KP090946      1113 TGCCCTCCTCACCCACGGCATCTCAGCGGAGATGCCATTCCGCGCCTTAATGCCTCGGC
GCQL01011182 1529 -----
GAGW01003914 288 -----
LKHE01001673 5921 -----
ANOV01001719 20730 -----
JAAVMX010000006 5597936 -----
LWBQ01000045 386808 -----

```

Figure S12: Alignment of the mRNA sequence KP090946 of *H. sinensis* strain L0106 with the genome and transcriptome sequences. KP090946 is the ribose-phosphate pyrophosphokinase-like mRNA sequence of *H. sinensis* strain L0106 [33]. LKHE01001673, ANOV01001719, JAAVMX010000006, and LWBQ01000045 are the genome assembly sequences of *H. sinensis* strains 1229, Co18, IOZ07, and ZJB12195, respectively [26,28,35-36]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.

```

KP090949          332 TTCAGATCCTCGAGGTCCTGGAATATCACGTCGTCTGCGTTGATGTGCTTGGCAATGTCC
GCQL01009009     1283 -----
LKHE01001105     8115 -----
LWBQ01000048     425162 -----
ANOV01001694     9761 -----
JAAVMX010000003 15195889 -----

KP090949          392 TGTCTCGTCTT                               GCCGTGC
GCQL01009009     1343 -----
LKHE01001105     8055 -----GCCGTGCGCGATGAGCTCGTTGGGGCTGGCGAGGTCAAT-----TAG
LWBQ01000048     425102 -----GCCGTGCGCGATGAGCTCGTTGGGGCTGGCGAGGTCAAT-----TAG
ANOV01001694     9701 -----GCCGTGCGCGATGAGCTCGTTGGGGCTGGCGAGGTCAAT-----TAG
JAAVMX010000003 15195829 -----GCCGTGCGCGATGAGCTCGTTGGGGCTGGCGAGGTCAAT-----TAG

KP090949          410 GCGATGAGC   TCGTTGGGGCTGG   CGAGGTCA ATGCCGTGA   ATGTGC
GCQL01009009     1361 -----
LKHE01001105     7995 GC-----T---ATCG--A-----CA--TTT-C--T--CG-G---CG--CTCACA-----
LWBQ01000048     425042 GC-----T---ATCG--A-----CA--TTT-C--T--CG-G---CG--CTCACA-----
ANOV01001694     9641 GC-----T---ATCG--A-----CA--TTT-C--T--CG-G---CG--CTCACA-----
JAAVMX010000003 15195769 GC-----T---ATCG--A-----CA--TTT-C--T--CG-G---CG--CTCACA-----

KP090949          455 GGATACGTGATGGGAGGCGCGCAGCTCGCAAAGATGACCTTGCGGGCCCCAGCCTCCCTC
GCQL01009009     1406 -----
LKHE01001105     7935 -----G-----
LWBQ01000048     424982 -----G-----
ANOV01001694     9581 -----G-----
JAAVMX010000003 15195709 -----G-----

KP090949          515 GCCATGCTGACAATCTCTCTGCTTGTCTGCGTCCCGGAACGATGGAGTCATCAACCAAGCAT
GCQL01009009     1466 -----
LKHE01001105     7875 -----
LWBQ01000048     424922 -----
ANOV01001694     9521 -----
JAAVMX010000003 15195649 -----

```

Figure S13: Alignment of the mRNA sequence KP090949 of *H. sinensis* strain L0106 with the assembled transcriptome and genome sequences. The mRNA sequence KP090949 and the assembled transcriptome sequence GCQL01009009 were obtained from *H. sinensis* strain L0106 [33]. LKHE01001105, LWBQ01000048, JAAVMX010000003, and ANOV01001694 are the genome assembly sequences of *H. sinensis* strains 1229, ZJB12195, IOZ07, and Co18, respectively [26,28,33,35-36]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.

```

KP090959          945  GCCCATTGGCTGGACGGCCATGCCGCTCGACGCTCGCTTCAGCACCGTGC GCGTCAAGGA
GCQL01011621     1312 -----
GAGW01006965     1052 -----
LWBQ01000131     207574 -----
LKHE01002574     90040 -----
ANOV01001374     16172 -----
JAAVMX01000005  9372625 -----A-----

KP090959          1005 ATCCAACATTGGCAACTTTGTCTGCGACGTGATGCGACGCTACCACAGCGCCGACTGCAC
GCQL01011621     1252 -----
GAGW01006965     1112 -----T-----
LWBQ01000131     207634 G--A-GG-A-----
LKHE01002574     89980 -----
ANOV01001374     16112 -----
JAAVMX01000005  9372565 -----

KP090959          1185 CCAAGCCATTTGGGATGCTTTGGAGAATGGTGTTCGTTGTATCCCGCTCTAGAGGGCAG
GCQL01011621     1072 -----
GAGW01006965     1292 -----
LWBQ01000131     207814 -----
LKHE01002574     89800 -----C-----
ANOV01001374     15932 -----C-----
JAAVMX01000005  9372385 -----

KP090959          1245 GTTCTCTCACGTCTCCAACATGGCCTACGAGTTCGACCCGAGCAGGGAGAGTGGC AAGA
GCQL01011621     1012 -----
GAGW01006965     1352 -----
LWBQ01000131     207874 -----A-----T
LKHE01002574     89740 -----
ANOV01001374     15872 -----
JAAVMX01000005  9372325 -----

KP090959          1304 GACTCACCTCCTTGTGCATCGGGGGCGAGGAATGGATCCCGGAGAAGCAGTATCTTCTCG
GCQL01011621     953 -----
GAGW01006965     1411 -----
LWBQ01000131     207934 -G- A-----
LKHE01002574     89681 -----
ANOV01001374     15814 -----
JAAVMX01000005  9372266 -----

KP090959          1364 CGACGAGAGGATACATGGGGCGCGGAAAAG          ACGGCTT CACCAGCTTG
GCQL01011621     893 -----
GAGW01006965     1471 -----
LWBQ01000131     207993 -----GTGAAAACAACC---T---GTG-G---GA
LKHE01002574     89621 -----GTGAAAACAACC---T---GTG-G---GA
ANOV01001374     15753 -----GTGAAAACAACC---T---GTG-G---GA
JAAVMX01000005  9372206 -----GTGAAAACAACC---T---GTG-G---GA

KP090959          1411 CTCGTGAA          AT          CAGAGGGC
GCQL01011621     846 -----
GAGW01006965     1518 -----
LWBQ01000131     208053 -G-----GCTGACGAG--GTATG---C---TTCACCAGCTTGCTCGTGAAATCAGAGG
LKHE01002574     89561 -G-----GCTGACGAG--GTATG---C---TTCACCAGCTTGCTCGTGAAATCAGAGG
ANOV01001374     15693 -G-----GCTGACGAG--GTATG---C---TTCACCAGCTTGCTCGTGAAATCAGAGG
JAAVMX01000005  9372146 -G-----GCTGACGAG--GTATG---C---TTCACCAGCTTGCTCGTGAAATCAGAGG

KP090959          1429 GGCCAGGCGGAGGAACTTGTTCGATGAAGAGCACGGCATTCTGATTCAGCCATGTTGC
GCQL01011621     828 -----
GAGW01006965     1536 -----
LWBQ01000131     208113 GC-----
LKHE01002574     89500 GC-----
ANOV01001374     15631 GC-----
JAAVMX01000005  9372086 GC-----

```

Figure S14: Alignment of the mRNA sequence KP090959 of *H. sinensis* strain L0106 with the assembled transcriptome and genome sequences. The KP090959 mRNA sequence and the assembled GCQL01011621 transcriptome sequence were obtained from *H. sinensis* strain L0106 [33]. The transcriptome assembly sequence GAGW01006965 was obtained from natural *C. sinensis* [53]. The genome assembly sequences LWBQ01000131, JAAVMX01000005, LKHE01002574, and ANOV01001374 were obtained from *H. sinensis* strains ZJB12195, IOZ07, 1229, and Co18, respectively [26,28,35-36]. Hyphens indicate identical bases, and spaces denote unmatched sequence gaps.