Unique Features of a Japanese ‘Candidatus Liberibacter asiaticus’ Strain Revealed by Whole Genome Sequencing

Hiroshi Katoh¹, Shin-ichi Miyata³, Hiromitsu Inoue², Toru Iwanami¹

¹ NARO Institute of Fruit Tree Science, Tsukuba, Ibaraki, Japan, ² Kuchinotsu Citrus Research Station, NARO Institute of Fruit Tree Science, Minami-shimabara, Nagasaki, Japan

Abstract

Citrus greening (huanglongbing) is the most destructive disease of citrus worldwide. It is spread by citrus psyllids and is associated with phloem-limited bacteria in the genus ‘Candidatus Liberibacter’ are associated with greening. ‘Ca. L. africanus’ is mainly present in Africa [2]; ‘Ca. L. americanus’ is found in Brazil [3]. A third species, ‘Ca. L. asiaticus’ is particularly widespread in Asian countries as well as in Sao Paulo, Brazil and Florida, USA. ‘Ca. L. asiaticus’ is transmitted by phloem-feeding insect vectors, the Asian citrus psyllid Diaphorina citri [4] and the African citrus psyllid Trioza erytreae [5]. A new Liberibacter species, ‘Ca. L. solanacearum’, was recently associated with the emerging ‘zebra chip’ disease of potatoes in the U.S. and tomatoes in New Zealand [6].

Little is known about the genetic diversity of ‘Ca. L. asiaticus’; the bacteria are difficult to culture, although some successes have been reported [7,8,9]. Diversity studies of ‘Ca. L. asiaticus’ have been restricted to the 16S/23S rRNA genes, the omp gene region, the rplK–rplB–rpoBC, nusG–rplK operon sequence, or bacteriophage-type DNA polymerase region (DNA pol) [10–20]. However, the complete genomc sequence of the pathogenic ‘Ca. L. asiaticus’ Floridian strain ‘psy62’ (1.23 Mb) [21] has been determined, thus enabling genome-wide analysis. In fact, Chen et al. characterized variation in ‘Ca. L. asiaticus’ strains by using one repeat unit (AGAGACA) [22]. From the whole-genome sequence, we selected 25 simple sequence repeat loci, including one repeat unit reported by Chen et al. [22] and successfully differentiated ‘Ca. L. asiaticus’ strains using these SSR loci [23,24]. Zhou et al. identified two hypervariable genes in the prophage regions of the psy62 genome [25]. Morgan et al. improved real-time PCR detection of ‘Ca. L. asiaticus’ from citrus and psyllid hosts by using the prophage gene [26]. The whole-genome sequencing of ‘Ca. L. asiaticus’ Floridian psy62 strain significantly advanced the study of diversity in this species.

Zhang et al. [27] reported two highly related, circular bacteriophage-type genes associated with ‘Ca. L. asiaticus’, named SC1 and SC2. Both were found integrated into the ‘Ca. L. asiaticus’ Floridian UF506 strain genome as prophages [27]. SC1 was apparently a fully functional, temperate phage with a lytic cycle that was seemingly activated when its host bacterium was present in plants but not when in psyllids [27]. SC2 replicates as an excision plasmid when its ‘Ca. L. asiaticus’ host is present in either plants or psyllids [27]. These findings suggest the bacteriophage-type genes are important for infection and virulence expression. However, most of the Japanese ‘Ca. L. asiaticus’ strains lack the bacteriophage-type DNA polymerase gene [18,19]. In Floridian UF506, the bacteriophage-type DNA polymerase gene is flanked by SC1 and SC2. Thus, absence of the bacteriophage-type DNA polymerase gene in Japanese strains suggests they also lack SC1 and SC2. Thus, the Japanese strains have unique genomic features.

In contrast to ‘Ca. L. asiaticus’ Floridian strains psy62 and UF506, the whole genome sequence of a Japanese ‘Ca. L. asiaticus’ strain lacking the bacteriophage-type DNA polymerase gene has not been reported. Recently, the complete genome
sequence of the Chinese ‘Ca. L. asiaticus’ strains gxpsy [28] and A4 [29] were reported, although the latter remains in the draft form. Both Chinese ‘Ca. L. asiaticus’ strains also contained the bacteriophage-type DNA polymerase gene. The results encouraged us to perform whole-genome sequencing of a Japanese strain lacking this gene. Duan et al. [21] obtained a complete circular ‘Ca. L. asiaticus’ Floridian psy62 strain genome by metagenomic analysis of DNA extracted from a single ‘Ca. L. asiaticus’-infected psyllid. We used a similar method to obtain the complete genome of the uncultured ‘Ca. L. asiaticus’ Japanese strain Ishi-1.

Materials and Methods

Bacterial strains

Japanese ‘Ca. L. asiaticus’ strain Ishi-1 was used throughout the study. The strain was originally found in local citrus of unidentified cultivars on Ishigaki Island, Okinawa prefecture, Japan. The infected scion was sent to the NARO Institute of Fruit Tree Science (NIFTS) with permission from the plant quarantine office of Japan, and kept in the isolated greenhouse after grafting on rough lemon (Citrus jambhiri Lush) rootstocks. The strain Ishi-1 induced severe symptoms on rough lemon, yuzu (Citrus junos Tanaka, Figure 1) and other citrus cultivars.

Psyllid treatment

All experiments using live individuals of D. citri were performed in insect-proof growth chambers at 25°C with a 16L:8D photoperiod at the Kuchinotsu Citrus Research Station, NIFTS (Otsu 954, Kuchinotsu, Minamishimabara, Nagasaki 859-2501, Japan). Healthy fifth instars of psyllids were transferred to an HLB-affected rough lemon tree (Citrus jambhiri, approximately 40 cm in height) with a high titer of ‘Ca. L. asiaticus’ bacteria. After acquisition feeding for 20 days on the infected plant, nine emerged adults were reared individually for 20 days on healthy Citrus junos seedlings for incubating the HLB bacteria, and they were stored at -25°C.

DNA extraction and quantitative real-time PCR

Total DNA was purified from the entire body of single psyllids using the DNeasy Blood and Tissue Kit (Qiagen, Tokyo, Japan) and a plastic homogenizer pestle (As One, Tokyo, Japan).

Figure 1. Foliar symptoms on Yuzu (Citrus junos Tanaka) induced by ‘Ca. L. asiaticus’ Japanese Ishi-1. Severe yellowing on the leaves of a Yuzu plant kept in a closed chamber at the NARO Institute of Fruit Tree Science.
doi:10.1371/journal.pone.0106109.g001
according to the manufacturer's instructions, and eluted in 150 μL.

Individual psyllid DNA samples were analyzed for *Ca. L. asiaticus* populations by quantitative real-time PCR analysis as described by Inoue et al. [30]. Samples containing copies of *Ca. L. asiaticus* genomic DNA were selected by real-time PCR (data not shown). Whole-genome amplification was performed with Illustra GenomiPhi V2 (GE Healthcare, Buckinghamshire, England) according to the manufacturer's instructions. DNA concentration was estimated with the Qubit 2.0 instrument and the Qubit dsDNA HS Assay (Life Technologies, Invitrogen, California).

Genome sequencing and mapping

Sequencing was performed at the Bio Dragon Genomics Center (Takara Bio Co. Ltd. Mie, Japan). DNA libraries with 300~350 bp inserts were constructed according to manufacturer's instructions (Illumina GaIIx platform) and 75-bp paired-end reads were generated on an Illumina HiSeqTM 2000 platform. Reads were mapped to the *Ca. L. asiaticus* Floridian psy62 genome using BWA [31] and Bowtie [32]. Mapping results were visualized with Integrative Genomics Viewer (IGV) version 2.3 [33].

Polymerase chain reaction for whole genome mapping confirmation

After initial genome mapping results were obtained, ambiguous sequences were determined by PCR amplification and conventional sequencing on an ABI 3130xl instrument. Total DNA was extracted from the leaf midrib tissue of citrus trees infected with the *Ca. L. asiaticus* Japanese Ishi-1 strain. Total DNA was extracted with the DNeasy plant minikit (Qiagen, Valencia, CA) according to manufacturer's instructions with minor modifications: approximately 0.2 g of the leaf midrib was placed in 400 μL AP1 buffer in a mortar and ground with a pestle until the leaf midrib became a fine green liquid.

Many In/Dels and SNPs were found by mapping the sequence reads of Ishi-1 to the complete sequence of the pathogenic *Ca. L. asiaticus* Floridian psy62 (1.23 Mb) strain, and primers were designed from the surrounding sequences (Primer3, http://frodo.wi.mit.edu/primer3/) (Table S1). Other primers were selected from Duan et al. [21]. PCR was performed with the Gene Amp PCR System 9700 (Applied Biosystems, Foster City, CA) in 20-μl reactions containing 1 μl DNA template, 0.1 μM each primer, 200 μM dNTPs, 1× PCR buffer, and 2.5 units of Ex Taq DNA polymerase Hot Start Version (TaKaRa, Shiga, Japan) under the following cycling conditions: initial denaturation at 92°C for
2 min; 35 cycles of denaturation at 92°C for 30 s, annealing at 54°C for 30 s, and extension for 1 min/kb of the desired product at 72°C. Long-range PCR of products above 3.0 kbp was performed with Tks Gflex DNA polymerase (TaKaRa). Each 50-μl reaction contained 1 μl DNA template, 0.1 μM each primer, 2× Gflex PCR buffer (Mg²⁺, dNTP plus), and 1.25 U Tks Gflex DNA polymerase. Cycling conditions were as follows: 30 cycles of denaturation at 98°C for 10 s, annealing at 52°C for 15 s, and extension for 30 s/kb of the desired product at 68°C.

DNA sequences were aligned using GENETYX-windows ver. 11 (Software Development, Tokyo, Japan), and homology analysis was performed as recommended by the DNA Data Bank of Japan (http://www.ddbj.nig.ac.jp/Welcome-j.html).

Gene prediction and functional annotation

Gene prediction and functional annotation were performed with the Microbial Genome Annotation Pipeline (MiGAP) (http://www.migap.org/index.php/en, [34]). Detection of tRNAs and rRNA was performed with tRNAScan-SE 1.23 (http://lowelab.ucsc.edu/tRNAscan-SE/, [35]) and RNAmmer 1.2 (http://www.cbs.dtu.dk/services/RNAmmer/, [36]).

Results

Whole genome re-sequencing

Sequencing yielded 2,721,927,150 bp of DNA from 36,292,362 pair-end reads of 75 bp. As a result of mapping using BWA, Bowtie against the psy62 strain reference, reads of 14.6% of the 2.7 Gbp were mapped. Coverage to the reference was 96.9%. The sequence reads of Ishi-1 were not mapped near the nucleotide position of 0 to 7,803 and after nucleotide position 1,195,171 of the linear genomic map of psy62, indicating that Ishi-1 lacks a large genomic fragment. PCR amplification of the Ishi-1 template with the LJ754r and LJ764f primers, which are separated by about 35 kbp on the psy62 sequence, yielded a 2.6-kbp product. Sequence analysis showed that the 2.6 kbp fragment filled the 35-kbp gap in the Ishi-1 genome.

Table 1. Comparison of the whole genome among three strains of ‘Ca. Liberibacter asiaticus’ and ‘Ca. L. solanacearum’.

| Features          | ‘Ca. L. asiaticus’ | ‘Ca. L. solanacearum’ |
|-------------------|-------------------|-----------------------|
|                   | Ishi-1            | psy62                 | gppsy                 | Clso-ZC1            |
| Size(bp)          | 1,190,853         | 1,227,328             | 1,268,237             | 1,258,278           |
| GC%               | 36.3              | 36.5                  | 36.5                  | 35.2                |
| rRNA operons      | 13                | 9                     | 6                     | 9                   |
| tRNA              | 44                | 44                    | 44                    | 45                  |
| RBS               | 975               | 1022                  | 1078                  | 1093                |
| CDS               | 1075              | 1134                  | 1165                  | 1192                |
| hypothetical protein | 313            | 358                   | 368                   | 409                 |

Table 2. Base substitution, insertion, deletion mutation and repeat number at respective SSR motif between ‘Ca. L. asiaticus’ Japanese Ishi-1 and Floridian psy62 strain.

|                         | Ishi-1 | psy62 | SSR motif | Ishi-1 | psyche |
|-------------------------|--------|-------|-----------|--------|--------|
| Base substitution       | 291    | –     | TACAGAA   | 14     | 8      |
| Insertion (one base insertion) | 66 (58) | –     | AGACACA   | 8      | 5      |
| Deletion (One base deletion) | 56 (38) | –     | TTG       | 9      | 14     |
|                         |        |       | TTATT     | 5      | 3      |
|                         |        |       | AGA       | 6      | 5      |

General features of the ‘Ca. L. asiaticus’ Japanese Ishi-1 genome

The calculated GC content of the Ishi-1 genome is 36.32%, similar to other ‘Ca. L. asiaticus’ strains (Table 1). After annotation, the newly confirmed CDS regions were compared to those of other ‘Ca. L. asiaticus’ strains. Then, the tRNAs and rRNA of Japanese Ishi-1 were compared to those of Floridian
Table 3. List of CDS encoded in the large 33 kbp fragment that is retained by Floridian psy62, UF506 and Chinese gxpsy strains, but not by Japanese Ishi-1 strain of ‘Ca. L. asiaticus.’

| psy62/UF506/gxpsy | product | locus tag | Locus location in genomea | product | locus tag | Locus location in genomeb | product | locus tag | Locus location in genomenc |
|-------------------|---------|-----------|--------------------------|---------|-----------|--------------------------|---------|-----------|--------------------------|
| hypot. protein    | CLIBASIA_00005 | 36..407 | hypot. protein | SC1_gp185 SC2_gp185 | 7649..76820 36401..36772 | hypot. protein | WSI_05545 | 1214938..1215309 |
| hypot. protein    | CLIBASIA_00010 | 497..820 | hypot. protein | SC1_gp190 SC2_gp190 | 76910..77233 36862..37185 | hypot. protein | WSI_05550 SC1_gp195 SC2_gp195 | 77361..78527 37313..38479 | hypot. protein | WSI_05555 SC1_gp195 SC2_gp195 | 1215850..1217016 1255717..1256883 |
| prophage          | CLIBASIA_00020 | 2285..3073 | putative Bro-N family phage antirepressor | SC1_gp200 SC2_gp200 | 78698..79486 32107..32292 | prophage antirepressor | WSI_05560 SC1_gp200 SC2_gp200 | 78698..79486 32107..32292 | prophage antirepressor | WSI_05560 SC1_gp200 SC2_gp200 | 1217187..1217978 1257054..1257845 |
| hypot. protein    | CLIBASIA_00025 | 3091..3741 | hypot. protein | SC1_gp205 SC2_gp205 | 79504..80154 39456..40106 | hypot. protein | WSI_05565 SC1_gp205 SC2_gp205 | 79504..80154 39456..40106 | hypot. protein | WSI_05565 SC1_gp205 SC2_gp205 | 1217996..1218646 1257863..1258513 |
| putative DNA poly. | CLIBASIA_00030 | 3745..5772 | DNA polymerase A | SC1_gp210 SC2_gp210 | 80158..812185 40110..42137 | putative DNA polymerase from bacteriophage origin | WSI_05570 SC1_gp210 SC2_gp210 | 80158..812185 40110..42137 | putative DNA polymerase from bacteriophage origin | WSI_05570 SC1_gp210 SC2_gp210 | 1218650..1220677 1258517..1260544 |
| VRR-NUC domain-    | CLIBASIA_00035 | 5769..6080 | endonuclease | SC1_gp215 SC2_gp215 | 82182..82493 42134..42445 | VRR-NUC domain-containing protein | WSI_05575 SC1_gp215 SC2_gp215 | 82182..82493 42134..42445 | VRR-NUC domain-containing protein | WSI_05575 SC1_gp215 SC2_gp215 | 1220674..1220985 1260541..1260852 |
| hypot. protein    | CLIBASIA_00040 | 6065..6727 | SNF2 Dead box helicase | SC2_gp220 | 42430..43815 | SNF2 related protein | WSI_05580 SC2_gp220 | 42430..43815 | SNF2 related protein | WSI_05580 SC2_gp220 | 1220970..1223255 1260837..1262222 |
| DNA ligase, NAD-  | CLIBASIA_00050 | 7442..7801 | DNA ligase | SC2_gp225 SC1_gp225 | 4811..5170 43808..44167 | DNA ligase, NAD-dependent | WSI_05360 SC2_gp225 SC1_gp225 | 4811..5170 43808..44167 | DNA ligase, NAD-dependent | WSI_05360 SC2_gp225 SC1_gp225 | 1183651..1184010 1222348..1222707 |
| guanylate kinase  | CLIBASIA_05525 | 1195911..1196264 | hypot. protein | SC1_gp235 | 44890..45261 | guanylate kinase | WSI_05595 SC1_gp235 | 44890..45261 | guanylate kinase | WSI_05595 SC1_gp235 | 1223448..1223780 |
| hypot. protein    | CLIBASIA_05530 | complement(1196268..1196741) | hypot. protein | SC1_gp005 | complement(45265..45738) | hypot. protein | WSI_05600 complement(45265..45738) | WSI_05600 complement(45265..45738) | hypot. protein | WSI_05600 complement(45265..45738) | 1223784..1224257 |
| hypot. protein    | CLIBASIA_05531 | complement(1196738..1197010) | hypot. protein | SC1_gp010 | complement(45735..46007) | hypot. protein | WSI_05610 complement(45735..46007) | WSI_05610 complement(45735..46007) | hypot. protein | WSI_05610 complement(45735..46007) | 1188796..1189068 |
| hypot. protein    | CLIBASIA_05538 | complement(1197003..1199762) | hypot. protein | SC1_gp025 | complement(46325..48448) | hypot. protein | WSI_05610 complement(46325..48448) | WSI_05610 complement(46325..48448) | hypot. protein | WSI_05610 complement(46325..48448) | 1224519..1226903 |
| hypot. protein    | CLIBASIA_05545 | complement(1199769..1202363) | hypot. protein | SC1_gp030 | complement(48455..51049) | hypot. protein | WSI_05615 complement(48455..51049) | WSI_05615 complement(48455..51049) | hypot. protein | WSI_05615 complement(48455..51049) | 1226910..1229504 |
| hypot. protein    | CLIBASIA_05550 | complement(1202360..1203796) | hypot. protein | SC1_gp035 | complement(51046..52482) | hypot. protein | WSI_05620 complement(51046..52482) | WSI_05620 complement(51046..52482) | hypot. protein | WSI_05620 complement(51046..52482) | 1229501..1230937 |
| hypot. protein    | CLIBASIA_05555 | complement(1203814..1205937) | hypot. protein | SC1_gp045 | complement(52500..54623) | hypot. protein | WSI_05625 complement(52500..54623) | WSI_05625 complement(52500..54623) | hypot. protein | WSI_05625 complement(52500..54623) | 1230955..1233078 |
| hypot. protein    | CLIBASIA_05560 | complement(1205934..1206449) | hypot. protein | SC1_gp050 | complement(54620..55123) | hypot. protein | WSI_05630 complement(54620..55123) | WSI_05630 complement(54620..55123) | hypot. protein | WSI_05630 complement(54620..55123) | 1233075..1233596 |
| psi62 product | locus tag       | Locus location in genome | UF506 product | locus tag       | Locus location in genome | gpxsy product | locus tag       | Locus location in genome |
|---------------|----------------|--------------------------|---------------|----------------|--------------------------|---------------|----------------|--------------------------|
| hypothetical protein | CLIBASIA_05565 | complement(120934..120949) | hypothetical protein | SC1_gp060 | complement(55116..59138) | hypothetical protein | WSI_05635 | complement(1233577..1237628) |
| hypothetical protein | CLIBASIA_05570 | complement(1210209..1210451) | hypothetical protein | SC1_gp080 | complement(59163..60899) | hypothetical protein | WSI_05640 | complement(1237645..1239381) |
| hypothetical protein | CLIBASIA_05575 | complement(1210476..1212212) | hypothetical protein | SC1_gp085 | complement(60892..61419) | hypothetical protein | WSI_05645 | complement(1239374..1239901) |
| hypothetical protein | CLIBASIA_05580 | complement(1212205..1212732) | hypothetical protein | SC1_gp090 | complement(61419..62450) | hypothetical protein | WSI_05455 | complement(1206042..1206173) |
| hypothetical protein | CLIBASIA_05585 | complement(1212732..1213763) | putative major capsid protein | SC2_gp090 | complement(22025..22945) | hypothetical protein | WSI_05565 | complement(1240893..1241594) |
| hypothetical protein | CLIBASIA_05590 | complement(1213776..1214480) | hypothetical protein | SC1_gp095 | complement(62463..63167) | hypothetical protein | WSI_05565 | complement(1240893..1241594) |
| hypothetical protein | CLIBASIA_05595 | complement(1214491..1214820) | hypothetical protein | SC1_gp100 | complement(63178..63507) | hypothetical protein | WSI_05566 | complement(1241605..1241934) |
| head-to-tail joining protein, putative | CLIBASIA_05600 | complement(1214813..1216483) | putative phage-related head-to-tail joining protein | SC1_gp105 | complement(63500..65170) | head-to-tail joining protein, putative | WSI_05565 | complement(1241927..1243597) |
| hypothetical protein | CLIBASIA_05605 | complement(1216480..1216812) | hypothetical protein | SC1_gp110 | complement(65167..65499) | hypothetical protein | WSI_05567 | complement(1243594..1243926) |
| putative phage terminase, large subunit | CLIBASIA_05610 | complement(1216885..1218420) | putative phage terminase, large subunit | SC1_gp115 | complement(65572..67107) | putative phage terminase, large subunit | WSI_05568 | complement(1244335..1245870) |
| hypothetical protein | CLIBASIA_05615 | complement(1218677..1218793) | hypothetical protein | SC1_gp120 | complement(67642..68157) | hypothetical protein | WSI_05569 | complement(1246405..1246893) |
| hypothetical protein | CLIBASIA_05620 | complement(1218955..1219443) | hypothetical protein | SC1_gp120 | complement(67642..68157) | hypothetical protein | WSI_05569 | complement(1246405..1246893) |
| hypothetical protein | CLIBASIA_05625 | complement(1220547..1221164) | hypothetical protein | SC1_gp125 | complement(69236..69853) | hypothetical protein | WSI_05569 | complement(1247997..1248614) |
| hypothetical protein | CLIBASIA_05630 | 1221334..1221936 | hypothetical protein | SC1_gp130 | 70104..70625 | hypothetical protein | WSI_05700 | complement(1248784..1249386) |
| hypothetical protein | CLIBASIA_05635 | 1221998..1222390 | hypothetical protein | SC1_gp135 | 70687..71079 | hypothetical protein | WSI_05500 | 1209271..1209663 | 1249448..1249840 |
| hypothetical protein | CLIBASIA_05640 | 1222526..1222732 | hypothetical protein | SC1_gp140 | 71215..71421 | hypothetical protein | WSI_05505 | 1209800..1210306 | 1249977..1250183 |
| hypothetical protein | CLIBASIA_05645 | 1222725..1222937 | hypothetical protein | SC1_gp145 | 71414..71626 | hypothetical protein | WSI_05510 | 1209999..1210211 | 1250176..1250388 |
| interrupted gp229 & phage associated protein | CLIBASIA_05650 | 1222969..1223301 | hypothetical protein | SC1_gp150 | 71658..72062 | hypothetical protein | WSI_05515 | 1210243..1210575 | 1250420..1250752 |
| hypothetical protein | CLIBASIA_05655 | complement(1223555..1223866) | hypothetical protein | SC1_gp160 | complement(72591..72968) | hypothetical protein | WSI_05725 | complement(1251006..1251317) |
## Table 3. Cont.

| product | psi62 | UF506 | GXpsy |
|---------|-------|-------|-------|
| P4 family phage/plasmid primase | CLIBASIA_05660 & complement(1223914..1226283) | P4 family phage/plasmid primase | CLIBASIA_05665 & complement(1226284..1226673) |
| SC1_gp165 & complement(73016..75388) | hypothetical protein | hypothetical protein |
| WSI_05730 & complement(1251365..1253734) | hypothetical protein | hypothetical protein |
| plasmid primase | primase | primase |
| associated | primase | primase |
| hypothetical protein | CLIBASIA_05665 & complement(1226284..1226673) | hypothetical protein | WSI_05535 & complement(1213869..1214258) |
| SC1_gp170 & complement(75389..75778) | hypothetical protein | hypothetical protein |
| WSI_05735 & complement(1253735..1254124) | hypothetical protein | hypothetical protein |
| SC2_gp170 & complement(35321..35710) | hypothetical protein | hypothetical protein |
| SC2_gp180 & complement(35931..36194) | hypothetical protein | hypothetical protein |
| hypothetical protein | CLIBASIA_05670 & complement(1226691..1226897) | hypothetical protein | WSI_05540 & complement(1214276..1214482) |
| CLIBASIA_05675 & complement(1226691..1227157) | hypothetical protein | hypothetical protein |

aData are based on the genome sequence of *L. asiaticus* Floridian psy62 strain. The accession number is CP001677 [21].

bData are based on the genome sequence of *Ca. L. asiaticus* Floridian UF506 strain. The accession number is HQ377374 [27].

L. asiaticus* Chinese gxpsy strain. The accession number is CP004005 [28].

eUnderline revealed that the deduced amino acid sequences showed about 80% similarity to one of *L. asiaticus* Floridian psy62 strain on the far left.

doi:10.1371/journal.pone.0106109.t003

These analyses revealed 1,075 coding sequences and 975 ribosome binding sites. We also found 44 tRNA genes that were shared with the Floridian psy62 and Chinese gxpsy strains, as well as 13 rRNA operons and 313 hypothetical proteins (Table 1). Our comparison of *Ca. L. asiaticus* Japanese Ishi-1 and Floridian psy62 revealed 291 base substitutions (Table 2). We also confirmed 122 in/del loci (Table 2). The five SSR loci were also polymorphic between the two strains (Table 2).

### Bacteriophage-type polymerase and other genes

As described above, the biggest difference in Japanese Ishi-1 is the absence of the 33-kbp fragment (Figure 3A). In psy62, this fragment encodes 40 CDS, including the bacteriophage-type polymerase gene between the SC1 and SC2 genes in the prophage region (Table 3). Most of the 40 CDS are shared between psy62, UF506, and Gxpsy (Table 3). None of these 40 CDS, including the bacteriophage-type polymerase gene, were found elsewhere in the genome of Ishi-1. Thus, Ishi-1 lacks the bacteriophage-type DNA polymerase gene found in Floridian strains psy62 and UF506, and the Chinese gxpsy strain ([21], [27], [28], shown by a vertical red line in Figure 3A). Another bacteriophage-type DNA polymerase is encoded in the middle of the linear schematic representation of the psy62 genome (shown by a vertical yellow line in Figure 3A). This bacteriophage-type DNA polymerase is also encoded in the corresponding region of the Ishi-1 genome (Figure 3A). Thus, it became clear that Ishi-1 carries a single bacteriophage-type DNA polymerase gene, whereas psy62 has two. In contrast, Chinese gxpsy and Floridian UF506 carry three bacteriophage-type DNA polymerase genes (WSI_05345, 05570, 05770, UF506_015, SC1_gp210, SC2_gp210).

Absence of the 33 kbp-fragment means other genes are also missing from the Ishi-1 genome. For example, Ishi-1 carries two NAD-dependent DNA Ligase genes (CGUJ_05395, 05515), whereas psy62 carries three (CLIBASIA_00050, 05395, 05515)—as does Chinese gxpsy (WSI_05360, 05585, 05785). In addition, the putative phage terminase, large subunit, exists in a single copy (CGUJ_05470) in the genome of Ishi-1, but as two copies in psy62 (CLIBASIA_05470, 05475, 05680). Furthermore, the number of Ishi-1 does not contain a full-length P4 family phage/plasmid primase gene; psy62 carries one (CLIBASIA_05515)—as does Chinese gxpsy (WSI_05360, 05585, 05785). In addition, the putative phage terminase, large subunit, exists in a single copy (CGUJ_05470) in the genome of Ishi-1, but as two copies in psy62 (CLIBASIA_05470, 05475, 05680). Furthermore, the number of Ishi-1 does not contain a full-length P4 family phage/plasmid primase gene; psy62 carries one (CLIBASIA_05515)—as does Chinese gxpsy (WSI_05360, 05585, 05785). In contrast, Chinese gxpsy and Floridian UF506 carry three bacteriophage-type DNA polymerase genes (WSI_05345, 05570, 05770, UF506_015, SC1_gp210, SC2_gp210).

### Characteristics of ‘Ca. L. asiaticus’ Japanese Ishi-1 strain marked by large In/Del variations

Several large In/Dels are shown in the simplified schematic presentation of the genome (Figure 3A). The 147-bp deletion at nucleotide positions 507106 through 507252 of the Floridian psy62 strain was detected in the genome of ‘Ca. L. asiaticus’ Japanese Ishi-1 (Figure 3B). This deletion reduced the hypothetical protein sequence at CGUJ_03230 by 49 amino acids in comparison to CLIBASIA_03230 of psy62 and WSI_02190 of Chinese gxpsy (Figure 3B). In contrast, the 2,108 bp insertion between nucleotide positions 983990 and 983991 (Figure 3C), an untranslated region in psy62, was detected in Ishi-1. This insertion carries a prophage anti-repressor at CGUJ_04441, and two hypothetical proteins at CGUJ_04442 and CGUJ_04443 were newly confirmed. The deduced amino acid sequence of the prophage anti-repressor at CGUJ_04441 is identical to that of Chinese gxpsy (WSI_04270), but different from those of Floridian psy62 and UF506. The hypothetical protein at CGUJ_04442 was also identical to that of Chinese gxpsy (WSI_04275). The hypothetical protein at CGUJ_04443 locus shared 99% amino
acid sequence identity with the putative WSI_04280 in Chinese gxpsy. These two hypothetical proteins shared no identity with any of the hypothetical proteins from Floridian psy62 and UF506.

Another insertion around nucleotide position 1081791 (psy62) was detected in Ishi-1 (Figure 3D), encoding a hypothetical protein at CGUJ_04911 within the 1,660-bp span. The deduced amino acid sequence of the hypothetical protein shares 48% identity with the hypothetical protein CKC_03455 from \('Ca. L. solanacearum'\) CLso-ZC1, a pathogen of zebra chip. Ishi-1 also carries a 149 bp-long insertion that correspond to the nucleotide positions 1182471 and 1182472 of psy62 (Figure 3A). No reading frames were found in the insertion.

Other In/Del and non-synonymous SNPs affecting annotation of \('Ca. L. asiaticus' Japanese Ishi-1\)

Lin et al. noted the absence of a full-length N-acetylglutamate kinase (NAGK) in the genome of \('Ca. L. asiaticus' Floridian psy62\), although it is present in \('Ca. L. solanacearum'\) CLso-ZC1 [37]. However, Japanese Ishi-1 (CGUJ_01846) and Chinese gxpsy (WSI_00005) encode identical full-length NAGK. Within the three \('Ca. L. asiaticus'\) strains, psy62 lacks an adenine between 406695 and 406696, thus truncating the sequence. The presence of an NAGK coding sequence indicates that Ishi-1 has a complete pathway for the production of arginine from glutamine, unlike psy62 (Figure 4).

Because of a single base insertion, Ishi-1 has two copies of the malic enzyme gene at CGUJ_00080 and CGUJ_00081, while \('Ca. L. asiaticus' Floridian psy62 (CLIBASIA_00080) and Chinese gxpsy (WSI_00005) each carry a single copy.

The genome of Ishi-1 also encodes a non-heme ferritin-like protein (CGUJ_00080), just like psy62 [38,39]. This ferritin-like protein is also found in \('Ca. L. solanacearum'\) [37], but is absent from the genomes of all other Rhizobaceae. The ferritin superfamily of proteins includes several diverse members that are typically involved in iron storage and detoxification [40,41,42]. Lin et al. hypothesized that this ferritin-like protein may play a critical role in the survival and/or virulence of \('Ca. L. solanacearum' and \('Ca. L. asiaticus'\) [37]. The genome of \('Ca. L. asiaticus' Chinese gxpsy also encodes a ferritin-like protein (WSI_2370). These sequences were identical but for a single amino acid substitution in Japanese Ishi-1 (Figure S1). In contrast, Floridian UF506 does not encode ferritin, indicating that this protein is dispensable.

Lin et al. reported that the \('Ca. L. solanacearum'\) genome encodes three known proteins involved in DNA replication and repair, all of which are absent from \('Ca. L. asiaticus' Floridian strain.

### Table 4. Presence of deduced amino acid sequence related DNA replication.

|       | LexA |       | DNAE |       | RadC |
|-------|------|-------|------|-------|------|
| locus tag | nucleotide position | locus tag | nucleotide position | locus tag | nucleotide position |
| Ishi-1 | -    | -     | CGUJ_03631 | 789771..793445 | CGUJ_03976 | ccomplement(866926..867195) |
| psy62  | -    | -     | -     | -     | -    | 874850..875119 |
| gxpsy  | -    | -     | WSI_03515 | 782257..785931 | WSI_03815 | complement(859383..859676) |
| CLso-ZC1 | CKC_02355 | 507699..508370 | CKC_05200 | complement(111198..111978) | CKC_04675 | 986531..987244 |

---

[Link to the PLOS ONE article](https://doi.org/10.1371/journal.pone.0106109.g004)
psychy2: LexA, DnaE, and RadC [37]. The genome Japanese Ishi-1 does not encode LexA, but does encode DnaE at CGUJ_03631 and RadC at CGUJ_03976. Chinese gxpsy also encodes DnaE and RadC, at WSL_03515 and WSL_03015, respectively (Table 4). However, a nucleotide sequence identical to that encoding RadC on the other ‘Ca. L. asiaticus’ strains is also carried in psy62 (Table 4), while LexA and DnaE are absent. This difference might be an annotation error.

The hypothetical protein at CGUJ_03991 was newly confirmed because of a one-base substitution in the untranslated region of the psy62 genome. The hypothetical proteins shared no similarity to other proteins of ‘Ca. L. asiaticus’. These proteins are listed in Table S2.

Discussion

As described previously, ‘Ca. L. asiaticus’, Japanese Ishi-1 lacks a bacteriophage-type DNA polymerase gene [19]. Our study showed that Ishi-1 lacks a large fragment of about 33 kbp that contains the bacteriophage-type DNA polymerase gene. It is noteworthy that this strain is found only in Japan; despite having the smallest genome of all ‘Ca. L. asiaticus’ strains, Floridian UF506 carries the large 33-kbp fragment [27]. We suggest the large 33-kbp fragment is associated with neither pathogenicity nor transmissibility, because Ishi-1 induced severe symptoms on citrus and propagated to a high titer in the vector insect. This is in sharp contrast to the discussions of Zhang et al. [27] regarding UF506, where the SC1 and SC2 genes flanking the bacteriophage-type DNA polymerase gene are suspected to be important for infection and virulence expression. It is likely that Ishi-1 carries different virulence factors. Most Japanese strains also lack the bacterio-

phage-type DNA polymerase gene [19]. Thus, the large 33-kbp fragment encoding the bacteriophage-type DNA polymerase gene may be absent from other Japanese strains, although confirmation by sequencing is needed. Another bacteriophage-type DNA polymerase is encoded in the middle of the Floridian psy62 and Japanese Ishi-1 genomes (Figure 3A), while Chinese gxpsy and Floridian UF506 carry two additional polymerases. These differences also suggest Ishi-1 (and perhaps other Japanese strains) are distinct from other strains from the US and China.

Zhou et al. identified two related and hypervariable genes (hyv-I and hyv-II) in the large 33-kbp fragment of the psy62 genome [25]. Although all DNA samples were obtained from symptomatic tissue and tested positive by 16S rRNA gene-based real-time PCR, neither the hyv-I nor the hyv-II gene was amplified from eight Indian citrus DNA samples and six Philippine psyllid DNA samples using the same primer sets [25]. These 14 strains likely lack the large 33-kbp fragment as Japanese Ishi-1 does. Thus, ‘Ca. L. asiaticus’ lacking the large fragment are not limited to Japan but are widespread in South Asia, pending confirmation by genome sequencing. We conclude it is best not to use primer sets specific to the large 33-kbp fragment for PCR detection of ‘Ca. L. asiaticus’, because some strains might escape detection.

Two malic enzyme genes were identified in the genome sequence of Ishi-1 in contrast to ‘Ca. L. asiaticus’ Floridian psy62 and Chinese gxpsy. The malic enzyme of the microaerophilic protist Entamoeba histolytica decarboxylates malate to pyruvate [43]. In ‘Ca. L. asiaticus’, a phloem-limited bacterium [2], the enzyme might play a similar role to that of E. histolytica. However, both malic enzyme genes of Ishi-1 are shorter than those of other ‘Ca. L. asiaticus’ strains. The Ishi-1 enzymes might not maintain their original function, suggesting a possible limitation of the Ishi-1 fermentation pathway.

In conclusion, whole genome sequencing of Japanese ‘Ca. L. asiaticus’ strain Ishi-1 revealed unique genomic features and suggested novel expression of virulence and establishment in host plant as well as distinct molecular evolution. We hope this study will advance our understanding of ‘Ca. L. asiaticus’ and facilitate efforts to control this devastating disease in the citrus industry.

Supporting Information

Figure S1 Comparison of deduced amino acid sequences of ferroxidase from ‘Ca. L. asiaticus’. Ferroxidase sequences were aligned by CLUSTAL W [48], and identical residues are indicated with asterisks. Databank accession numbers are CP001677 for psy62 [21], AP014595 for Ishi-1, and CP004005 for gxpsy [28], respectively. (TIFF)

Table S1 Conventional and long-distance PCR primers used for ‘Ca. L. asiaticus’ Japanese Ishi-1 strain sequence confirmation and gap closure in this study. (XLSX)

Table S2 List of CDS in the genome of ‘Candidatus Liberibacter asiaticus’ Japanese Ishi-1 strain that revealed no deduced amino acid sequence similarity to those of other ‘Ca. L. asiaticus’ strains. (XLSX)

Acknowledgments

We thank Ms. H. Hatomi for providing meticulous technical assistance to this work.

Author Contributions

Conceived and designed the experiments: TI. Performed the experiments: HK SM HI. Analyzed the data: HK SM. Contributed reagents/materials/analysis tools: SM HI TI. Contributed to the writing of the manuscript: HK SM HI TI.

References

1. Zhao X (1981) Citrus yellow shoot disease (Huanglongbing)—A review. Proc Int Soc Citriculture 1: 466–469.
2. Jagoueix S, Bove JM, Garnier M (1994) The phloem-limited bacterium of greening disease of citrus is a member of the β subdivision of the Proteobacteria. Int J Syst Bacteriol 44: 379–386.
3. Teixeira D (2005) First report of a Huanglongbing-like disease of citrus in Sa˜o Paulo State, Brazil, and association of a new Liberibacter species, ’Candidatus Liberibacter asiaticus’, with the disease. Plant Dis 89: 107.
4. Halbert SE, Manjunath KL (2004) Asian citrus psyllids (Sternorrhyncha: Psyllidae) and greening disease of citrus: a literature review and assessment of risk in Florida. Fla Entomol 87: 330–353.
5. Bove JM (2006) Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. J Plant Pathol 88: 7–37.
6. Lifting LW, Perez-Egasquina ZC, Clover GRG, Anderson JAD (2008) A new ’Candidatus Liberibacter’ species in Solanum tuberosum in New Zealand. Plant Dis 92: 1474.
7. Garnett HM (1984) Isolation of the greening organism. Citrus Subtrop Fruit J 61: 4–5.
8. Davis MJ, Mondal SN, Chen H, Rogers ME, Belansky RH (2008) Co-cultivation of ‘Candidatus Liberibacter asiaticus’ with Actinobacteria from citrus with Huanglongbing. Plant Dis 92: 1547–1550.
9. Secher A, Schuenzel EL, Cooke P, Donnua S, Thaveechai N, et al. (2009) Cultivation of ‘Candidatus Liberibacter asiaticus’, ‘Ca. L. africanus’, and ‘Ca. L. americanus’ associated with Huanglongbing. Phytopathology 99: 480–486.
10. Bastianel C, Garnier-Semancik M, Renaudin J, Bove JM, Euvrard S (2005) Diversity of ‘Candidatus Liberibacter asiaticus’, based on the omp gene sequence. Appl Environ Microbiol 71: 6473–6478.
11. Ding F, Deng X, Hong N, Zhou Y, Wang G, et al. (2009) Phylogenetic analysis of the citrus Huanglongbing (HLB) bacterium based on the sequences of 16S rDNA and 16S/23S rDNA intergenic regions among isolates in China. Eur J Plant Pathol 124: 493–503.

12. Furuya N, Matsukura K, Tomimura K, Okuda M, Miyata S, et al. (2010) Sequence homogeneity of the \textit{spc\textsubscript{E}rpl} gene cluster and the flanking regions of \textit{Candidatus Liberibacter asiaticus} isolates around Okinawa Main Island in Japan. J Gen Plant Pathol 76: 122–131.

13. Furuya N, Tomimura K, Okuda M, Miyata S, et al. (2011) Recombination-like sequences in the upstream region of the phase-type DNA polymerase in \textit{Candidatus Liberibacter asiaticus}. J Gen Plant Pathol 77: 295–298.

14. Jagoueix S, Bové JM, Garnier M (1997) Comparison of the 16S/23S ribosomal gene clusters and the flanking regions of \textit{Candidatus Liberibacter asiaticus} isolates around Okinawa Main Island in Japan. J Gen Plant Pathol 76: 122–131.

15. Miyata S, Kato H, Davis R, Smith MW, Weinert M, et al. (2011) Asian common strain of \textit{Candidatus Liberibacter asiaticus} are distributed in Northeast India, Papua New Guinea and Timor-Leste. J Gen Plant Pathol 77: 43–47.

16. Villechanoux S, Garnier M, Laigret F, Renaudin J, Bové JM (1993) The genome of the non-cultured, bacterial-like organism associated with citrus greening disease contains the \textit{nuG}
\textit{rpKJA}\textsubscript{L}\textsubscript{pOBC} gene cluster and the gene for a bacteriophage type DNA polymerase. Curr Microbiol 30: 137–141.

17. Subandiyah S, Iwanami T, Yasuyuki S, Ieki H (2000) Comparison of 16S rDNA and 16S/23S intergenic region sequences among isolates in China. Plant Dis 84: 15–18.

18. Duan YL, Zhou DG, Hall W, Li H, Doddapaneni H, et al. (2009) Complete genome sequence of \textit{Candidatus Liberibacter asiaticus} obtained through metagenomics. Mol Plant-Microbe Interact 22: 1011–1020.

19. Chen J, Deng X, Sun X, Jones D, Irey M, et al. (2010) Guangdong and Florida populations of \textit{Candidatus Liberibacter asiaticus} distinguished by a genomic locus with short tandem repeats. Phytopathology 100: 567–572.

20. Katoh H, Subandiyah S, Tomimura K, Okuda M, Su HJ, et al. (2011) Differentiation of \textit{Candidatus Liberibacter asiaticus} isolates collected in Southeast Asia. Phytopathology 99: 1062–1069.

21. Tomimura K, Furuya N, Miyata S, Hamashima A, Torigoe H, et al. (2010) Distribution of Two Distinct Genotypes of Citrus Greening Organism in the Ryukyu Islands of Japan. Jpn Agric Res Q 44: 151–158.

22. Chen J, Deng X, Sun X, Jones D, Irey M, et al. (2010) Guangdong and Florida populations of \textit{Candidatus Liberibacter asiaticus} distinguished by a genomic locus with short tandem repeats. Phytopathology 100: 567–572.

23. Katoh H, Subandiyah S, Tomimura K, Okuda M, Su HJ, et al. (2011) Differentiation of \textit{Candidatus Liberibacter asiaticus} isolates collected in Southeast Asia. Phytopathology 99: 1062–1069.

24. Katoh H, Davis R, Smith MW, Weinert M, Iwanami T (2012) Differentiation of \textit{Candidatus Liberibacter asiaticus''} isolates isolated in Southeast Asia. Phytopathology 99: 1062–1069.

25. Tomimura K, Furuya N, Miyata S, Hamashima A, Torigoe H, et al. (2010) Distribution of Two Distinct Genotypes of Citrus Greening Organism in the Ryukyu Islands of Japan. Jpn Agric Res Q 44: 151–158.

26. Villechanoux S, Garnier M, Laigret F, Renaudin J, Bové JM (1993) The genome of the non-cultured, bacterial-like organism associated with citrus greening disease contains the \textit{nuG}
\textit{rpKJA}\textsubscript{L}\textsubscript{pOBC} gene cluster and the gene for a bacteriophage type DNA polymerase. Curr Microbiol 30: 137–141.

27. Subandiyah S, Iwanami T, Yasuyuki S, Ieki H (2000) Comparison of 16S rDNA and 16S/23S intergenic region sequences among isolates in China. Plant Dis 84: 15–18.

28. Zheng Z, Deng X, Chen J (2014) Whole-Genome Sequence of \textit{Candidatus Liberibacter asiaticus} from Guangdong, China. Genome Announc 2 (2), e00275-14.

29. Inoue H, Ohnishi J, Ito T, Tomimura K, Miyata S, et al. (2009) Enhanced proliferation and efficient transmission of \textit{Candidatus Liberibacter asiaticus} by adult \textit{Diaphorina citri} after acquisition feeding in the nymphal stage. Ann Appl Biol 155: 29–36.

30. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1734–1740.

31. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1734–1740.

32. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

33. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, et al. (2011) Integrative genomics viewer. Nat Biotech 29: 24–26.

34. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl Acids Res 25: 3100–3108.

35. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl Acids Res 25: 3100–3108.

36. Lagesen K, Hallin P, Rodland EA, Stærfeldt H-H, Rognes T, et al. (2007) RACEAmr: consistent and rapid annotation of ribosomal RNA genes. Nucl Acids Res 35: 3100–3108.

37. Lin H, Lou B, Glynn JM, Doddapaneni H, Civerolo EL, et al. (2011) The complete genome sequence of \textit{Candidatus Liberibacter solanacearum}, the bacterium associated with potato zebra chip disease. PLoS ONE 6: E19135.

38. Reindel S, Anemuller S, Savaryn A, Marzanne BF (2002) The DpsA homologue of the \textit{archaeon} \textit{Halobacterium salinarum} is a ferritin. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 1598: 140–146.

39. Zeth K, Ollermann S, Esen L-O, Oesterhelt D (2004) Iron-oxo clusters biominingarization on protein surfaces: Structural analysis of \textit{Halobacterium salinarum} DpsA in its low- and high-iron states. Prox Nac Acad Sci USA 101: 13780–13783.

40. Andrews SC (2010) The ferritin-like superfamily: Evolution of the biological iron storeman from a ruberythrin-like ancestor. Biochimica et Biophysica Acta (BBA) - General Subjects 1800: 691–705.

41. Carvendo MA (2003) Ferritins, iron uptake and storage from the bacterioferritin viewpoint. EMBO J 22: 1959–1968.

42. Zhao G, Ceci P, Ilari A, Giangiacomo L, Laue TM, et al. (2002) Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells: A ferritin-like DNA-binding protein of \textit{Escherichia coli}. J Biol Chemistry 277: 27689–27696.

43. Field J, Rosenhal B, Samuelson J (2000) Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to \textit{Entamoeba histolytica}. Mol Microbiol 38: 446–455.

44. Fernández-Murga ML, Gil-Oritz F, Lácer JL, Rubio V (2004) Arginine biosynthesis in \textit{Thermotoga maritima}: Characterization of the Arginine-Sensitive N-Acetyl-L-Glutamate Kinase. J Bacteriol 186: 6142–6149.

45. Hass D, Holloway BW, Schambrock A, Leuzinger T (1977) The genetic organization of arginine biosynthesis in \textit{Pseudomonas aeruginosa}. Mol Gen Genet 154: 7–22.

46. Ibeda M, Minehashi S, Tanaka K, Hayashi M (2009) Reengineering of a \textit{Corynebacterium glutamicum} \textit{L-Arginine} and \textit{L-Citrulline} Producer. Appl Environ Microbiol 75: 1635–1641.

47. Pécard FJ, Dillon JR (1989) Cloning and organization of seven arginine biosynthesis genes from \textit{Neisseria gonorrhoeae}. J Bacteriol 171: 1644–1651.

48. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res 22:4673–4680.