groE-Homologous Operon of Wolbachia, an Intracellular Symbiont of Arthropods: A New Approach for Their Phylogeny

Authors: Masui, Shinji, Sasaki, Tetsuhiko, and Ishikawa, Hajime

groE-Homologous Operon of *Wolbachia*, an Intracellular Symbiont of Arthropods: A New Approach for Their Phylogeny

Shinji Masui, Tetsuhiko Sasaki and Hajime Ishikawa*

Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

ABSTRACT—*Wolbachia*, a member of rickettsia found in the cells of many arthropod species, are cytoplasmically inherited bacteria which interfere with host's sexuality and reproduction. *Wolbachia* strains have been phylogenetically divided into A and B groups based on the nucleotide sequences of their *ftsZ* genes. In an attempt to further define the phylogenetical relationship among these endosymbionts, we cloned and sequenced the entire length of the *groE* operon of a *Wolbachia* harbored by a cricket. The operon encoded two heat shock proteins, which represented the third and fourth proteins of any *Wolbachia* ever characterized. Also, 800 bp stretches of the *groE* operons of several other *Wolbachia* were sequenced, and a phylogenetic tree was constructed based on the results. The *groE* tree defined the relationship among A group *Wolbachia* strains that had not been successfully resolved by the *ftsZ* tree, and suggested unexpected horizontal transmission of these bacteria.

INTRODUCTION

Wolbachia, a rickettsia-like microorganism, is present in the cells of many species of arthropod, mainly in insects (O'Neill *et al.*, 1992; Juchault *et al.*, 1994; Werren *et al.*, 1995). These endosymbiotic bacteria are transmitted maternally through the host egg, and known to alter host reproduction in various ways including post-zygotic reproductive incompatibility, or cytoplasmic incompatibility (CI), in a wide range of insects (Barr, 1980; O'Neill *et al.*, 1992; Breeuwer *et al.*, 1992), parthenogenesis in wasps (Stouthamer *et al.*, 1993), and feminization of genetic male in an isopod (Rousset *et al.*, 1992). Molecular mechanisms underlying these phenomena are still unknown.

To date, except for the rRNA genes, only two protein coding genes from *Wolbachia* were sequenced, which are *ftsZ* involved in regulation of bacterial cell division, and *dnaA* essential for initiation of DNA replication (Holden *et al.*, 1993; Bourtzis *et al.*, 1994). Phylogenetic trees of *Wolbachia* strains in the various hosts have been constructed based on the sequences of 16S rRNA and parts of the coding region of *ftsZ* (O'Neill *et al.*, 1992; Werren *et al.*, 1995; Tsagkarakou *et al.*, 1996). The *ftsZ* tree had finer resolution than that of 16S rRNA, and divided *Wolbachia* into two major groups designated as A and B. It was suggested that the two had diverged from each other 58-67 million years ago (Werren *et al.*, 1995). The tree also implied frequent horizontal transmissions of

Wolbachia between distantly related insect orders. In A-group of *Wolbachia*, horizontal transmission takes place so frequently that there is increasing need for a less conserved sequence than that of *ftsZ* to understand detailed phylogeny and infection pathways of this clade.

The *groE* operon, encoding highly conserved bacterial heat shock protein GroES and GroEL which are also called HSP10 and HSP60, respectively, contains the noncoding, intergenic region between these two coding regions (Hartl, 1996; Segal and Ron, 1996). The intergenic sequence is almost completely neutral to natural selection, and, thus, believed to evolve faster than the coding sequences. The organization of the *groE* operons is highly conserved, and it is known that there is only one copy of the *groE* operon per genome in the *Rickettsiaceae* group to which *Wolbachia* strains belong (Segal and Ron, 1996; O'Neill *et al.*, 1992; Roux and Raoult, 1995). Thus, the intergenic region of *groE* operon is a suitable material based on which we construct a phylogenetic tree of closely related *Wolbachia* strains.

In this study, we sequenced *groE*-homologous operons of *Wolbachia* including that from an infected cricket, *Teleogryllus taiwanemma*, and compared their sequences among several strains of A-group *Wolbachia* whose relationship has been quite ambiguous. As a result, we were successful in defining phylogenetic relationship among these *Wolbachia* strains with higher resolution than previously reported.

---

* Corresponding author: Tel. +81-3-5800-3553;
             FAX. +81-3-5800-3553.

MATERIALS AND METHODS

Insect materials

Taiwan crickets, *T. taiwanemma*, were reared at 24°C under the 16 hr light and 8 hr dark conditions. They were fed on artificial diet, CA-1 (CLEA JAPAN), and tap water. To eliminate *Wolbachia*, crickets were given 0.25% (w/v) tetracycline hydrochloride at least for 3 generations. Moths of sub-family of *Phycitinae*, *Ephestia kuehniella*, *Ephestia cautella* and *Plodia interpunctella* were reared on wheat bran containing 10% (w/w) glycerol at 24°C with a photoperiod of 16 hr. An aposymbiotic strain of *E. kuehniella* was established by rearing the insects on a diet containing 0.04% (w/w) tetracycline hydrochloride for two generations. *Drosophila simulans* Hawaii (DSH) and *Drosophila simulans* Riverside (DSR) were provided by Dr. O'Neill. DNA was extracted using DNAzol™ reagent (GIBCO BRL) from dissected ovaries of *T. taiwanemma*. As for other insects, DNA was extracted from the whole bodies.

PCR amplification and sequencing

To amplify part of the *groE* sequence of *Wolbachia*, degenerate primers were designed on the basis of an alignment of *Ehrlichia chaffeensis* (GenBank accession no. L10917), *Cowdria ruminantium* (U13638) and *Orientia* (*Rickettsia*) *tsutsugamushi* (M31887) *groEL* gene sequences. The degenerate primers were: WgLf 5'-TGANGAAG-AAATTGCNCAAGT-3' (*E. chaffeensis groEL* positions 417-437), WgLr 5'-CCTTCTTCAACTGCAGCTCTTGN-3' (1235-1213).

To clone its flanking regions, cassette PCRs (LA PCR *in vitro* Cloning Kit, TAKARA) were performed according to manufacturer's recommendations. Bands obtained were cloned into pGEM-T (Promega) and sequenced with T7 and SP6 primers using an automated sequencer (SQ-5500, HITACHI). To eliminate PCR errors, at least 3 clones were sequenced, or PCR products were directly sequenced. Primers designed for amplification of 800 bp fragments encompassing the intergenic region of the *groE* operon were: groEfl 5'-TGTATTAGATGATAACGTGC-3' (*Wolbachia groE* operon positions 21-40), groErl 5'-CCATTTGCAGAAATTATTGCA-3' (844-824). A-group specific primers of this region were also designed: groEAf 5'-TGATCAAGCCTATTAGC-3' (41-57), groEAr 5'-GAGATTATTGCAA-CTTGTGCC-3' (835-815). PCR conditions when these primers were employed were 30 cycles with 94°C 1 min, 52°C 1 min, 72°C 2 min.

The sequence analyzed in this study have been deposited in GenBank under accession numbers AB002286 for the full length of *Wolbachia groE* operon from *T. taiwanemma*, and AB002287-91 for the partial sequences of *groE* operons from other *Wolbachia* strains.

Phylogenetic analysis

CLUSTAL W (Thompson *et al.*, 1994) was used to align the sequences, to construct the NJ tree (including gap positions, with correction for multiple substitution), and to calculate bootstrapping probabilities (1000 resamplings). The program MEGA (ver. 1.02, Kumar *et al.*, 1993) was used to estimate number of nucleotide substitutions per site (d_A) and number of nonsynonymous and synonymous substitution per nonsynonymous (d_N) and synonymous (d_S) site, respectively, using Jukes-Cantor correction and excluding insertions-deletions.

RESULTS

groE operon of *Wolbachia*

We first amplified and cloned part of the *groEL*-homologous gene of *Wolbachia* harbored by a cricket, *T. taiwanemma*. This cricket is relatively large in size, and the ovary could be isolated easily without contamination by enterobacteria. PCR using degenerate primers reproducibly amplified a 0.8 kb fragment, which was shown to hybridize with the genomic DNA

from the ovary by Southern blot analysis (data not shown). Secondly, the complete nucleotide sequence of the *groE* operon was determined (Fig. 1) by cassette PCRs based on the sequence of the 0.8 kb fragment obtained above. The intergenic region between the *groES* and *groEL* gene consisted of 90 bp, the length of which was typical of *groE* operons of *Rickettsiaceae*. The operon contained two ORFs that encoded 96 amino acids (GroES) and 552 amino acids (GroEL) with molecular mass of 10471 and 58965 Da, respectively. The operon contained typical ribosome-binding-sites (Stormo, 1986) with GGAA or GGAG poly-purine stretch.

The neighbour-joining analysis of the deduced *Wolbachia* GroEL amino acid sequence was performed (figure not shown), which indicated that *Wolbachia* clearly formed a clade with *E. chaffeensis* (86 out of 100 bootstrap replications). Congruence of the tree topology with that of the 16S rRNA tree (O'Neill *et al.*, 1992; Roux and Raoult, 1995) confirmed that this gene is from *Wolbachia*.

For phylogenetic analysis of *Wolbachia* strains, we chose a 800 bp region in the *groE* operon which included the entire sequence of the intergenic region and its flanking sequences (see Fig. 1), considering that these regions are highly susceptible to base substitutions. To amplify the 800 bp *groES-L* region, *groE* general primers were designed (Fig. 2), and the specificity of these primers was tested by PCR assay (Fig. 3). As a result, positive signals were detected from infected insects, but not from a naturally uninfected species and tetracycline-cured individuals, again confirming that this gene is from *Wolbachia*. It has been shown from the *ftsZ* gene analysis that *T. taiwanemma* contains B-group *Wolbachia* (our unpublished data), while *E. cautella* harbors both A and B groups (Werren *et al.*, 1995; Furukawa, unpublished data) which were designated as *E. cautella* A and *E. cautella* B, respectively. DSH, DSR, and *E. kuehniella* contain A-group (Werren *et al.*, 1995; Furukawa, unpublished data). Using *groE* general primers, we amplified and sequenced the *groES-L* region of *Wolbachia* from *E. kuehniella*, which was used to design A-group specific *groE* primers, groEAf and Ar (Fig. 2). PCR assay showed that these primers successfully amplified 800 bp fragment corresponding to the *groES-L* region from insects known to have A-group *Wolbachia*, but not from the one that contained B-group (see Figs. 2 and 3).

Comparison of the *groES-L* regions

The numbers of substitutions per site (d_A) between the *groES-L* regions of the two representatives of A and B-group *Wolbachia*, from *E. kuehniella* and *T. taiwanemma*, respectively, were estimated. The d_A of the intergenic sequence was 0.16 ± 0.048, which was similar to the average d_A of the coding region, 0.16 ± 0.012. However, it should be emphasized that the intergenic region contained many insertions and deletions (13 bp insertions and 2 bp deletions in total, based on *T. taiwanemma*) which were not considered here in estimating the d_A.

For the GroES and GroEL coding sequences, d_N and d_S were separately calculated. The values for d_N were very low

```
AAGCTTATACCATTTATCTTATCGATTAAATATTGCAGCGGAAATGATGAATTTTGCACCATATCTTTGCTCCATCTGGGACAAGGTTTTTAATTCTAACTTTTTCATACTGAATAAAA   -540

TCTCCAACCGGTTCTTTTATTTCATCTTCATGTATGATCCTCATTCCTAGCAGCTTGTGCCTTGGATCTTGAAAAATAACTTGTGACTTACTACTACACTTCGCCAACTTAGCATCAAAC   -420

AAAACTCCAACCTTATATAGTGAACTAATGTCTTTAATCCTCACTCTAAGGTAAGTTTTGAGCAAGTCTAATTTTTCAATTATTTGCTGTAGATGGGCATTTTCACACTCCAAATAAATA   -300

TATTTATCATGCTCAATGAGGAAAAAATCGTACAGATATTTTCCCTGAGGGTTGAGTAATAAAGAGTAAATTGCCTTCTGGCTACTTAGTTTATTAATATCATTTGTTATAACACCCTGA   -180

AGAAAATCTCTCGTATCTGGTCCGTATAAAGATATTAGACTACGGTTTGCTAGCGGTATGTAACCCATAAAACTTCTGTTCAAACCTTAATTATACACTTTTACTAACTCCAACCCTTGA    -60

AATGTAACCTTTAGATGACTACATAGAATTGTAGGTTAATCTATTAATGAGGAAACATAAATGTCAAGTGTAAATTTGAATGTATTAGATGATAACGTGCTAATAAAGCCTATTACTGAA     60
                                          M  S  S  V  N  L  N  V  L  D  D  N  V  L  I  K  P  I  T  E   20

GAAAAACAAGGTGGAATTGTACTACCATCAAGTGCTGAAAAGAAGCCTAAAAAAGGTGAAGTTATAGCAATTGGTAGTGGTTCACGCAACTCAAGCGGAGAGCGTATAGCTTTAACTGTA    180
 E  K  Q  G  G  I  V  L  P  S  S  A  E  K  K  P  N  K  G  E  V  I  A  I  G  S  G  S  R  N  S  S  G  E  R  I  A  L  T  V   60

AAAACTGGTGATAAGGTTTTCTACAGGCAGTGGGCTGGGACAGAGGTAGAGCATGATAATGAAAAGTATGTCGTGATGAAGGAGTCAGACCTACTTGCTGTTATCAAGTAGGTATTCCTA    300
 K  T  G  D  K  V  F  Y  R  Q  W  A  G  T  E  V  E  H  D  N  E  K  Y  V  V  M  K  E  S  D  L  L  A  V  I  K  *          96

TTTTACTAAACTAATGCGCTATATAGTGCATTATGTATTTTAATATTCTAAATTTTTTATTAATTTAAGGAGTGATAAAAATGGCTAACATAGTAGTATCAGGCGAACAGTTGCAAGAAG    420
         intergenic region                                            M  A  N  I  V  V  S  G  E  Q  L  Q  E   14

CCTTTCGTGAAGTTGCAGCAATGGTTGACTCAACGGTCGGGAGCAACTGCAGGGCCTAGGGGAAATACAGTAGGGATTAGCAAGCCATATGGTGGACCAGAGGTCACAAAAGATGGCTATA    540
 A  F  R  E  V  A  A  M  V  D  S  T  V  G  A  T  A  G  P  R  G  N  T  V  G  I  S  K  P  Y  G  G  P  E  V  T  K  D  G  Y   54

AGGTAATGAAAGGTATAAAGCCTGAAAAACCATTACACTCTGCGATAGTAAGCACCATTGCTCAAAGTGCTTCTCAGTGTAATGACAAAGTTGGTGATGGTACCACAACATGCTCAATAC    660
 K  V  M  K  G  I  K  P  E  K  P  L  H  S  A  I  V  S  T  I  A  Q  S  A  S  Q  C  N  D  K  V  G  D  G  T  T  T  C  S  I   94

TGACTAGCAATATGATAATGGAAGCTTCAAAGTCAATTGCAGCTGGAAATGATCGTATTTGTATCAAAAATGGAATACAGAAGGCAAAAGATGTGATATTAAAGGGAGATTACATCAATGT    780
 L  T  S  N  M  I  M  E  A  S  K  S  I  A  A  G  N  D  R  I  C  I  K  N  G  I  Q  K  A  K  D  V  I  L  K  E  I  T  S  M   134

CTCGCACAATTTCTTTAGAGAAAATGGATGAAGTTGCACAAGTTGCAATAATTTCTGCAAATGGTGATAAAGATATTGGTAATAGCATTGCTGATGCTGTTAAAAAGGTTGGAAAGAAG    900
 S  R  T  I  S  L  E  K  M  D  E  V  A  Q  V  A  I  I  S  A  N  G  D  K  D  I  G  N  S  I  A  D  A  V  K  K  V  G  K  E   174

GCGTCATCACTGTTGAAGAGAGTAAGGGTTCAAAAGAATTAGAAGTTGAACTTACAACTGGTATGCAGTTTGACCGTGGTTATCTCTCTCCATACTTTGTTACAAATAATGAAAAGATGA   1020
 G  V  I  T  V  E  E  S  K  G  S  K  E  L  E  V  E  L  T  T  G  M  Q  F  D  R  G  Y  L  S  P  Y  F  V  T  N  N  E  K  M   214

GCGTGGAGCTTGATGATCCATATTTGCTGATTACAGAAAAAAAACTTAATATTATTCAACCTTTACTTCCTATTCTTGAGGCTATTTTTAAGTCTGGTAAACCTTTGTTTATCATTGCAG   1140
 S  V  E  L  D  D  P  Y  L  L  I  T  E  K  K  L  N  I  I  Q  P  L  L  P  I  L  E  A  I  F  K  S  G  K  P  L  F  I  I  A   254

AAGATGTTGAAGGTGAAGCATTAAGCACTTTAGTGATCAACAAATTACGTGGTCTAAAAGTTGCTGCAGTAAAAGCTCCAGGTTTTGGTGATAGGAGAAAGGAGATGCTCGAAGATATAG   1260
 E  D  V  E  G  E  A  L  S  T  L  V  I  N  K  L  R  G  L  K  V  A  A  V  K  A  P  G  F  G  D  R  R  K  E  M  L  E  D  I   294

CAGCTTTAACAGGTGCTAAGTATGTCATAAAAGATGAGCTTGGAATCAAGATGGAAGATTTGACTCTTGAAGATCTTGGTACTGCTAAAAATGTTAAAATCACTAAAGACAATACTACAA   1380
 A  A  L  T  G  A  K  Y  V  I  K  D  E  L  G  I  K  M  E  D  L  T  L  E  D  L  G  T  A  K  N  V  K  I  T  K  D  N  T  T   334

TTGTTAGTGAAGACAGCGACTGTGACAAACAAAACAGAGTAAATGCTAGAATTAACCGATTAAATCTCAGATTGAAACTTCAACCTGTGATTATGATAAAGAAAAGCTGAGAGAGCGTT   1500
 I  V  S  E  D  S  D  C  D  K  Q  N  R  V  N  A  R  I  N  Q  I  K  S  Q  I  E  T  S  T  S  D  Y  D  K  E  K  L  R  E  R   374

TAGCGAAATTATCAGGCGGCGTTGCTGTGCTCAAAGTTGGTGGAGCAACCGAAGTAGAAGTAAAAGAACGCAGAGATAGAGTTGAAGATGCACTTCATGCAACAAGAGCTGCAATTGAAG   1620
 L  A  K  L  S  G  G  V  A  V  L  K  V  G  G  A  T  E  V  E  V  K  E  R  R  D  R  V  E  D  A  L  H  A  T  R  A  A  I  E   414

AAGGGATAGTTCCAGGTGGAGGAGTTGCACTTCTTTACGCTGCATCTGCTCTTGATAAGCTAAAAGCTTCAAGCGATGAAGAGCAAATAGGTATCAACATTGTCAAAAAAGTTCTCAGTG   1740
 E  G  I  V  P  G  G  G  V  A  L  L  Y  A  A  S  A  L  D  K  L  K  A  S  S  D  E  E  Q  I  G  I  N  I  V  K  K  V  L  S   454

CTCCAATCAAAAGATTGGTTAAAAATGCAGGTCTTGAATCTGCTGTTATAATTGACTATTTGATTAAGCAGAATGATAAAGAGCTTATATACAACGTTGAGGCTATGAACTACGCTAATG   1860
 A  P  I  K  R  L  V  K  N  A  G  L  E  S  A  V  I  I  D  Y  L  I  K  Q  N  D  K  E  L  I  Y  N  V  E  A  M  N  Y  A  N   494

CATCAACAGCTGGTGTAATTGATCCGGCAAAAGTAGTTCGTATTGCTTTTGAAACGGCAGTGTCTGTAGCAAGCGCACTTATTACTACTGAATCTATGATAGTTGATCTACCAAACAAAG   1980
 A  S  T  A  G  V  I  D  P  A  K  V  V  R  I  A  F  E  T  A  V  S  V  A  S  A  L  I  T  T  E  S  M  I  V  D  L  P  N  K   534

AAGAAAATGCTTCATCACCTATGGGTGCAGGAGCAATGGGCGGAATGGGTGGATTCTAAGTAAACAAGAGCCTTGGAGCAACTTGCTCCAAGGTATCTTTCTAGTTCTAAAAAGTCTTAC   2100
 E  E  N  A  S  S  P  M  G  A  G  A  M  G  G  M  G  G  F  *                                                           552

ATTTATAAAGGAAATAAATTTGTGATTTTCTATGCAAAGCCACTACTATTTATATTTGTTACGTCTCTTTAATATTATAACTGCTTAAAAATAAAAAAATAACATATGAAAGTTATAATT   2220

GCTAGTATATTTGTATTAGCATAATATTATATTAACAAATGATTTATATAATATTTTTTATTGTCCCAAATTTGGGAGAAAAACCCTGGACTTGAAGTCCATCCTTTTAAGAATATAATTT   2340

TAAGGTGAAAAATCTGAAAGAATGTGGAATATTATAGAAAAAGCAGTCACTCTGTATATGAATGTACATATCACATAGTGTGGATAACAAAATATCGATAGCGAATTCCTG           2451
```

Fig. 1. The complete nucleotide sequence of the *groE* operon of *Wolbachia* and the deduced amino acid sequence. Presumed ribosome-binding sites are marked by under lines. Asterisks represent stop codons. Upper and lower numbers indicate the position of the nucleotide sequence from the translation start site, and the amino acid positions of each protein, respectively. Arrowheads mark the 5' and 3' boundary of the region used for phylogenetic analysis in this study, which is also shown in Fig. 2.

and similar to that of *ftsZ* (data not shown). The d_S values for *groES* and *groEL* were 0.74 ± 0.18 and 0.58 ± 0.10, respectively, which were 1.3-1.6 fold higher than that of the *ftsZ* gene, 0.47 (Werren *et al.*, 1995). Similar d_S values were observed when the corresponding regions were compared between DSH and *E. cautella* B, 0.71 ± 0.18 for *groES* and 0.51 ± 0.096 for *groEL*.

Phylogenetic relationship

Phylogenetic relationship among 6 different *Wolbachia* strains was investigated by the Neighbour-Joining algorithm

(Saitou and Nei, 1987), using the 800 bp *groES-L* region including the intergenic sequence of the *groE* operon (Fig. 4). The average sequence divergence between A and B-group was 17.2%, after Jukes and Cantor correction. Within the A-group, the divergence ranged between 0.27 and 2.3%. While insertions-deletions were not detected in the 4 strains of the A-group used in this study, the *groE* tree, thus constructed, clearly defined the phylogenetic relationship among them. There were only 2 bp substitutions between the *groE* sequences from DSH and *E. cautella* A, which were the same substitution number between the *ftsZ* genes of these two
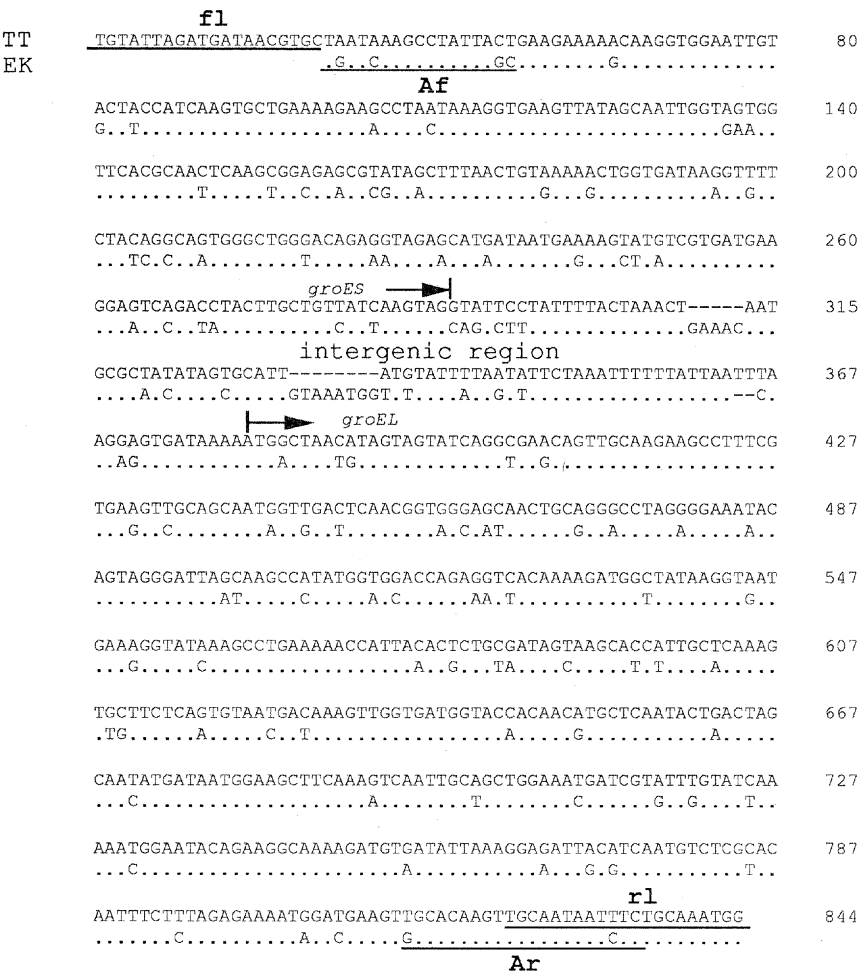
```
                        f1
        TT      TGTATTAGATGATAACGTGCTAATAAAGCCTATTACTGAAGAAAAACAAGGTGGAATTGT      80
        EK      .G..C.........GC........G..............
                        Af
        ACTACCATCAAGTGCTGAAAAGAAGCCTAATAAAGGTGAAGTTATAGCAATTGGTAGTGG      140
        G..T.................A....C...........................GAA..

        TTCACGCAACTCAAGCGGAGAGCGTATAGCTTTAACTGTAAAAACTGGTGATAAGGTTTT      200
        .........T.....T..C..A..CG..A...........G...G...........A..G..

        CTACAGGCAGTGGGCTGGGACAGAGGTAGAGCATGATAATGAAAAGTATGTCGTGATGAA      260
        ...TC.C..A........T.....AA....A...A........G...CT.A..........
                    groES        ►
        GGAGTCAGACCTACTTGCTGTTATCAAGTAGGTATTCCTATTTTACTAAACT-----AAT      315
        ...A..C..TA.........C..T......CAG.CTT..............GAAAC...
                        intergenic region
        GCGCTATATAGTGCATT--------ATGTATTTTAATATTCTAAATTTTTTATTAATTTA      367
        ....A.C....C.....GTAAATGGT.T....A...G.T..................--C.
                    ┠─►     groEL
        AGGAGTGATAAAAATGGCTAACATAGTAGTATCAGGCGAACAGTTGCAAGAAGCCTTTCG      427
        ..AG...........A....TG.............T..G..

        TGAAGTTGCAGCAATGGTTGACTCAACGGTGGGAGCAACTGCAGGGCCTAGGGGAAATAC      487
        ...G..C.........A..G..T.........A.C.AT......G..A.....A.....A..

        AGTAGGGATTAGCAAGCCATATGGTGGACCAGAGGTCACAAAAGATGGCTATAAGGTAAT      547
        ..........AT.....C.....A.C.......AA.T...........T........G..

        GAAAGGTATAAAGCCTGAAAAACCATTACACTCTGCGATAGTAAGCACCATTGCTCAAAG      607
        ...G.....C..................A..G...TA....C.....T.T....A.....

        TGCTTCTCAGTGTAATGACAAAGTTGGTGATGGTACCACAACATGCTCAATACTGACTAG      667
        .TG......A.....C..T.................A.....G...........A.....

        CAATATGATAATGGAAGCTTCAAAGTCAATTGCAGCTGGAAATGATCGTATTTGTATCAA      727
        ...C.....................A........T........C......G..G....T..
                                                rl
        AAATGGAATACAGAAGGCAAAAGATGTGATATTAAAGGAGATTACATCAATGTCTCGCAC      787
        ...C.....................A...........A...G.G...........T..

        AATTTCTTTAGAGAAAATGGATGAAGTTGCACAAGTTGCAATAATTTCTGCAAATGG       844
        .......C..........A..C.....G..................C...........
                                                Ar
```

Fig. 2. Locations of the *groE* general and A-group specific primers. Primer regions are underlined. The *groE* general, fl and rl primers were designed to amplify 800 bp *groES-L* sequences from A and B-group *Wolbachia*, while Af and Ar were designed specifically for A-group. The 3' terminal of *groES* and 5' terminal of *groEL* coding regions are indicated by arrows. Numbering shows the positions in the *groE* operon based on *T. taiwanemma* sequence (see Fig. 1). TT, *T. taiwanemma*; EK, *E. kuehniella*.

Wolbachia strains. In contrast, the *groE* sequences from *E. kuehniella* and *E. cautella* A, whose *ftsZ* genes are identical with each other (Furukawa, unpublished data), contained 17 bp substitutions, 5 bp of which were found in the intergenic region.
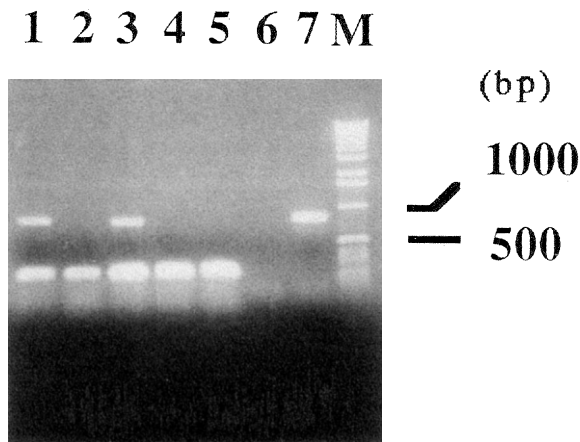

Fig. 3. PCR assay for the specificity of *groE* primers. The *groE* general and A-group specific primers were used in lanes 1-5 and 6-7, respectively (bands at 800 bp). Insect 18S rRNA was used in separate PCRs as control for quality of the DNA extraction (bands at 200 bp). In lanes 1-5, the PCR products were combined with reaction mixtures from the control PCRs. All samples were electrophoresed on 1.0% agarose gel, stained with ethidium bromide, and visualized with UV. Lanes: M, 1 kb DNA ladder (GIBCO BRL); 1, 2 and 6, *T. taiwanemma*; 3, 4 and 7, *E. kuehniella*; 5, *P. interpunctella*. Lanes 2 and 4; tetracycline-treated strains. *P. interpunctella* is a naturally uninfected insect.
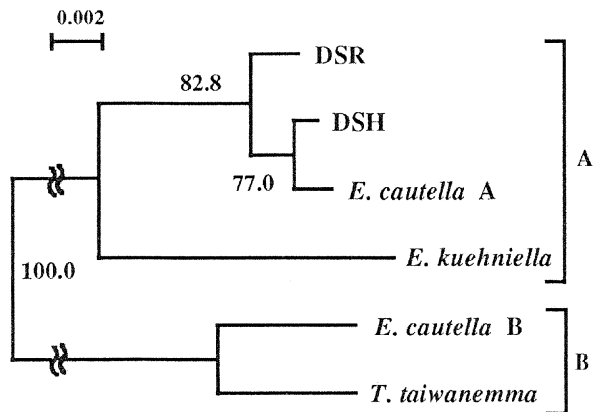
Fig. 4. Phylogenetic tree derived from the neighbour-joining analysis of the *groE* gene from 6 different *Wolbachia* strains. This unrooted tree was calculated by using CLUSTAL W, including positions of insertions-deletions and with correction for multiple substitutions. Numbers next to nodes indicate bootstrapping probabilities out of 1000.

DISCUSSION

The *groEL* gene of *Wolbachia*

The *groE* operons in eubacteria are highly conserved, and GroEL homologs from numerous eubacterial genera have been identified as the major antigens (Dasch *et al.*, 1990). In general, intracellular symbionts and parasites produce their GroEL homologs in large amounts (Choi *et al.*, 1991; Vodkin and Williams, 1988; Stover *et al.*, 1990). In an endosymbiont *Buchnera* of the pea aphid *Acyrthosiphon pisum*, a GroEL homolog called symbionin is selectively expressed (Hara *et al.*, 1990). A histidine residue at the position 133 of symbionin is prone to autophosphorylation (Morioka *et al.*, 1993), suggesting that this protein not only functions as molecular chaperone but also plays a role in signal transduction in this endosymbiotic bacterium (Gross *et al.*, 1989; Morioka *et al.*, 1994). In this study, we cloned and sequenced the *groE*-homologous operon of an endosymbiont *Wolbachia* from a cricket. The operon encoded the third and fourth proteins of *Wolbachia* that have been ever characterized. In *Wolbachia* GroEL, the His-133 of *Buchnera* GroEL had been replaced by a serine. Serine residue at this position was also observed in GroEL homologs of *E. chaffeensis* and *Arabidopsis* mitochondria (Swissprot accession no. P29197), which live in intracellular environment. It is conceivable that these serine residues also play a role in signal transduction in these bacteria and organelle through their phosphorylation.

Since the amino acid sequence of GroEL homologs is highly conserved, and their amino acid substitution is relatively free from the influence of biased base substitution typically known in the evolution of 16S rRNA (Hasegawa and Hashimoto, 1993; Jukes and Bhushan, 1986), it is widely used as an evolutionary chronometer (Viale, 1995). To date, the phylogenetic position of *Wolbachia* among alpha proteobacteria has been determined only by 16S rRNA (O'Neill *et al.*, 1992; Roux and Raoult, 1995). The neighbour-joining

analysis of the deduced amino acid sequence of *Wolbachia* GroEL supported the 16S rRNA tree, confirming that *Wolbachia* is positioned in the *Rickettsiaceae* family.

Wolbachia phylogeny by the *groES-L* sequences

Apparently recent spread of A-group *Wolbachia* among natural populations of *D. simulans* was reported (Turelli and Hoffmann, 1991), and it was considered to be due to human disturbance or transport (Werren *et al.*, 1995). In addition, the *ftsZ* phylogeny suggested that frequent horizontal transmission of A-group *Wolbachia* among various distantly related hosts including *Drosophila* and *Ephestia*. However, the infection pathway among them is unclear because their sequences determined have been almost identical with each other (Werren *et al.*, 1995).

It was shown that *groES-L* sequences analyzed in this study are very useful for phylogeny of A-group *Wolbachia*, because of its higher d_S values and the presence of the intergenic region where insertions-deletions can provide useful information. The topology of the *groE* tree and its high bootstrap values clearly elucidated relationship among the 4 strains of A-group *Wolbachia*, for which no divergence had been successfully indicated by the *ftsZ* phylogeny. In this tree, *E. cautella* A formed a clade with DSH rather than with *E. kuehniella* (Fig. 4). Since *E. cautella* and *E. kuehniella* are moths that share a similar nich, it is considered that horizontal transmission of *Wolbachia* occurs frequently between the two. However, the tree suggests that horizontal transmission has not occurred between the moths, but between moths and flies as the most recent event. Though the direction of the transmission is not clear with this number of host species examined here, it is obvious that the *groES-L* sequence is, potentially, a powerful tool to resolve phylogeny and infection pathways of *Wolbachia* much more precisely than before.

Wolbachia is rapidly spreading all over the world and infecting even nematodes (Siloni *et al.*, 1995), and seems to promote speciation of host by bringing incompatibility between host populations (Breeuwer and Werren, 1990; Coyne, 1992), suggesting that the *Wolbachia* infection can be a driving force of evolution. Further analysis of infection pathways among various hosts, will make clear the possible mechanisms of horizontal transmission, and the influences on recent speciation and evolution of host species.

REFERENCES

Barr AR (1980) Cytoplasmic incompatibility in natural populations of a mosquito, *Culex pipiens*. Nature 283: 71–72

Bourtzis K, Nirgianaki A, Onyango P, Savakis C (1994) A prokaryotic *dnaA* sequence in *Drosophila melanogaster. Wolbachia* infec-

tion and cytoplasmic incompatibility among laboratory strains. Insect Molec Biol 3: 131–142

Breeuwer JAJ, Werren JH (1990) Microorganisms associated with chromosome destruction and reproductive isolation between two insect species. Nature 346: 558–560

Breeuwer JAJ, Stouthamer R, Barns SM, Pelletier DA, Weisburg WG, Werren JH (1992) Phylogeny of cytoplasmic incompatibility microorganisms in the parasitoid wasp genus *Nasonia* (Hymenoptera: Pteromalidae) based on 16S ribosomal DNA sequences. Insect Molec Biol 1: 25–36

Choi EY, Ahn GS, Jeon KW (1991) Elevated levels of stress proteins associated with bacterial symbiosis in *Amoeba proteus* and soybean root nodule cells. BioSystems 25: 205–212

Coyne JA (1992) Genetics and speciation. Nature 355: 511–515

Dasch GA, Ching WM, Kim PY, Pham H, Stover CK, Oaks EV, Dobson ME, Weiss E (1990) A structural and immunological comparison of rickettsial HSP60 antigens with those of other species. Ann N Y Acad Sci 590: 352–369

Gross R, Arico B, Rappuoli R (1989) Families of bacterial signal-transducing proteins. Mol Microbiol 3: 1661–1667

Hara E, Fukatsu T, Kakeda K, Kengaku M, Ohtaka C, Ishikawa H (1990)The predominant protein in an aphid endosymbiont is homologous to an *E. coli* heat shock protein. Symbiosis 8: 271–283

Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381: 571–580

Hasegawa M, Hashimoto T (1993) Ribosomal RNA trees misleading? Nature 361: 23

Holden PR, Brookfield JFY, Jones P (1993) Cloning and characterization of an *ftsZ* homologue from a bacterial symbiont of *Drosophila melanogaster.* Molec Gen Genet 240: 213–220

Juchault P, Frecon M, Bouchon D, Rigaud T (1994) New evidence for femenizing bacteria in terrestrial isopods: evolutionary implications. CR Acad Sci Paris 317: 225–230

Jukes TH, Bhushan V (1986) Silent nucleotide substitutions and G+C content of some mitochondrial and bacterial genes. J Mol Evol 24: 39–44

Kumar S, Tamura K, Nei M (1993) MEGA: molecular evolutionary genetics analysis (Pennsylvania State Univ, University Park, PA), Version 1.02

Morioka M, Muraoka H, Ishikawa H (1993) Chaperonin produced by an intracellular symbiont is an energy-coupling protein with phosphotransferase activity. J Biochem 114: 246–250

Morioka M, Muraoka H, Yamamoto K, Ishikawa H (1994) An endosymbiont chaperonin is a novel type of histidine protein kinase. J Biochem 116: 1075–1081

O'Neill SL, Giordano R, Colbert AME, Karr TL, Robertson HM (1992) 16SrRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc Natl Acad Sci USA 89: 2699–2702

Rousset F, Bouchon D, Pintureau B, Juchault P, Solignac M (1992) *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. Proc R Soc Lond B 250: 91–98

Roux V, Raoult D (1995) Phylogenetic analysis of the genus *Rickettsia* by 16S rDNA sequencing. Res Microbiol 146: 385–396

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Molec Biol Evol 4: 406–452

Segal G, Ron EZ (1996) Regulation and organization of the *groE* and *dnaA* operons in Eubacteria. FEMS Microbiol Lett 138: 1–10

Siloni M, Bandi C, Sacchi L, Sacco BD, Damiani G, Genchi C (1995) Molecular evidence for a close relative of the arthropod endosymbiont *Wolbachia* in a filarial worm. Molec Biochem Parasitol 74: 223–227

Stormo GD (1986) Translation initiation. In "The Maximizing Gene Expression" Ed by W Reznikoff and L Gold, Butterworths, Boston, MA, pp 195–224

Stouthamer R, Breeuwer JAJ, Luck RF, Werren JH (1993) Molecular identification of microorganisms associated with parthenogenesis. Nature 361: 66–68

Stover CK, Marana DP, Dasch GA, Oaks EV (1990) Molecular cloning and sequence analysis of the Sta58 major antigen gene of *Rickettsia tsutsugamushi*: sequence homology and antigenic comparison of Sta58 to the 60-kilodalton family of stress proteins. Infect Immun 58: 1360–1368

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

Tsagkarakou A, Guillemaud T, Rousset F, Navajas M (1996) Molecular identification of a *Wolbachia* endosymbiont in a *Tetranychus urticae* strain (Acari: Tetranychidae). Insect Molec Biol 5: 217–221

Turelli M, Hoffmann AA (1991) Rapid spread of an incompatibility factor among natural *Drosophila simulans* populations. Nature 353: 440–442

Viale AM (1995) GroEL (HSP60)-based eubacterial and organellar phylogenies. Molec Microbiol 17: 1013

Vodkin MH, Williams JC (1988) A heat shock operon in *Coxiella burnetii* produces a major antigen homologous to a protein in both mycobacteria and *Escherichia coli*. J Bacteriol 170: 1227–1234

Werren JH, Zhang W, Guo LR (1995) Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. Proc R Soc Lond B 261: 55–71