Intervening Sequence Acquired by Lateral Gene Transfer in *Tropheryma whippiei* Results in 23S rRNA Fragmentation

Nicolas Crapoulet, Sylvianne Robineau, Didier Raoult, and Patricia Renesto*

Unité des Rickettsies, CNRS-UMR6020, IFR48, Faculté de Médecine, 27, Bd. Jean Moulin, Marseille, France

Received 23 February 2005/Accepted 19 July 2005

Completion of *Tropheryma whippiei* genome sequencing may provide insights into the evolution of the molecular mechanisms underlying the pathogenicity of this microorganism. The first postgenomic application was the successful design of a comprehensive culture medium that allows axenic growth of this bacterium, which is particularly recalcitrant to cultivation. This achievement in turn permitted analysis of *T. whippiei* RNA without contaminating eukaryotic nucleic acids. To obtain high-quality RNA, several extraction methods were compared, but under all conditions tested an atypical profile was observed. By using a Northern blot assay we demonstrated that an insertion sequence previously described in *T. whippiei* 23S rRNA is in fact an intervening sequence excised during maturation. This cleavage could involve an RNase III identified in the genome of this microorganism. Among the bacteria with a 23S rRNA insertion sequence, *T. whippiei* is the only gram-positive microorganism. We present phylogenetic evidence that this mobile genetic element was acquired by lateral gene transfer from another enteric bacterium.

Tropheryma whippiei is the gram-positive bacterium that is responsible for Whipple's disease (14). In the last few years, several steps that have improved our knowledge of this microorganism have been taken, starting from cultivation of the first human isolate in fibroblasts (20). This opened the way for *T. whippiei* genome sequencing (3, 21), which in turn could result in a better understanding of the molecular basis for the pathogenicity of this bacterium. The availability of the *T. whippiei* genome also allowed identification of several genes, including *rnc* encoding RNase III. RNase III specifically recognizes and cleaves 23S rRNA insertion sequences called intervening sequences (IVSs), a process originally identified in *Salmonella enterica* (4) and now described as a general feature of several bacteria, such as *Proteobacteria* (6) and *Spirochaetes* (19). Analysis of domain III of the *T. whippiei* 23S rRNA gene revealed the presence of an actinobacterial type B insertion sequence (accession no. AF148137) (10). In the present study, our aim was to look for possible excision of this insertion sequence, which is located at positions 1,541 to 1,621 of *rrl* in *T. whippiei* strain Twist. This investigation was made possible by another concrete postgenomic advance, namely, the design of a comprehensive culture medium that allows replication of *T. whippiei* under axenic conditions (22). Previously, cultivation of *T. whippiei* required the presence of eukaryotic host cells (7, 20). Given its small genome size, it is not surprising that *T. whippiei* lacks almost all known genes that are required for the de novo biosynthesis of amino acids. As is the case for other small-genome intracellular bacteria, *T. whippiei* depends on the availability of host-derived metabolites that are imported into the bacteria by a variety of transporters (1, 17). When RNA is extracted from obligate intracellular bacteria, such as *Rickettsia*

conorii, eukaryotic nucleic acids are also detected (24). This hampers fine analysis of the prokaryotic RNA profile. In this work, we took advantage of axenic cultivation of *T. whippiei* to determine if 23S rRNA insertion sequence cleavage occurs.

MATERIALS AND METHODS

Strain, medium, and growth conditions. All experiments were performed with mid-log cultures of *T. whippiei* strain Twist (12) grown under axenic conditions (22). Bacterial growth was monitored by flow cytometry counting using a Microcyte portable flow cytometer (Optoflow AS, Oslo, Norway) and by quantitative PCR as previously described (8, 22).

RNA extraction. All reagents were made and all processes were performed in sterile, disposable, guaranteed nuclease-free labware. All solutions were made with water treated with 0.1% (vol/vol) diethyl pyrocarbonate (DEPC). Following centrifugation (16,900 × g, 10 min) bacteria were resuspended in Trizol reagent (Invitrogen Life Technologies) and sonicated on ice for 0 s to 6 min (Vibra Cell 75022; Bioblock Scientific). RNA extraction was carried out according to the manufacturer's guidelines. The resulting pellet was resuspended in 40 µl of DEPC-treated H₂O and treated with DNA-free (Ambion, Austin, TX) to remove DNA contamination. The amount and integrity of each RNA sample were checked by automated capillary gel electrophoresis using a Bioanalyzer 2100 with RNA Nano LabChips (Agilent, Palo Alto, CA).

RT-PCR amplification. Reverse transcription (RT)-PCR amplification was performed using the one-step RT-PCR system (Superscript reverse transcriptase with Platinum *Taq* polymerase; Life Technologies). The reaction was carried out in a 25.2-µl mixture containing 12.5 µl of 2× reaction mixture, each primer at a concentration of 0.4 µM, 0.7 µl of RT/Platinum *Taq* mixture, and 2 µl of template RNA. The primers used were TwrpoB.F (5'-TTGAGCGCAGCCG GAAAAA-3') and TwrpoB.R (5'-GCACCGCAACCTCGGAGAAA-3'). These primers, which were designed to target a 507-bp fragment of the β-subunit of the RNA polymerase gene (*rpoB*), were demonstrated previously to be highly sensitive diagnostic tools (5). RT-PCR amplification was carried out with a PTC-100 thermocycler (MJ Research, Inc.). Negative controls containing 2 µl of recombinant *Taq* DNA polymerase (Gibco BRL) instead of 1 µl of the RT/Platinum *Taq* mixture were included in all reactions to determine whether there was DNA contamination. Amplified products were visualized with a UV transilluminator after electrophoresis on 1% (wt/vol) agarose gels and ethidium bromide staining.

Northern blot assay. Genomic *T. whippiei* DNA extracted using a QIAamp DNA mini kit (QIAGEN, Courtaboeuf, France) was amplified by PCR (Expand High Fidelity PCR system; Roche Applied Science, Indianapolis, IN) using primers amplifying total 23S rRNA (*rrl*) (2,326 bp; primers PIF [5'-CGTGGG GAAGTGAACATCT-3'] and PIR [5'-ACACTGCCACCTGATTACC-3']),

* Corresponding author. Mailing address: Unité des Rickettsies, CNRS UMR6020, IFR48, Faculté de Médecine, 27, Boulevard Jean Moulin, 13385 Marseille, France. Phone: (33) 491 32 43 75. Fax: (33) 491 38 77 72. E-mail: patricia.renesto@medecine.univ-mrs.fr.

as well as a 432-bp region located in the 5' 23S rRNA extremities (primers P2F [5'-CGTGGGGAAGTGAACATCT-3'] and P2R [5'-CCGACGGTTTGTAAGCAAAT-3']) or a 479-bp region located in the 3' 23S rRNA extremities (primers P3F [5'-AAGTGTCTCAACCGCAACT-3'] and P3R [5'-GGGAGGACCAACCTGTTAT-3']). The 16S rRNA (*rrs*) (403 bp; primers P4F [5'-CGCAAGGTCGGTATACAGGT-3'] and P4R [5'-CGGGTGTACCAGCTTTCAT-3']) and the 5S rRNA (101 bp; primers P5F [5'-GCAATGGCAAGAGGAAAC-3'] and P5R [5'-GCGTCTTACTCTCCACAGG-3']) were also amplified. The PCR products obtained were cloned into the pGEM-T Easy vector (Promega, Charbonnières, France), and positive sequenced clones (3100 genetic analyzer; Applied Biosystems, Courtaboeuf, France) were used as DNA templates for a second PCR amplification. Amplicons were radiolabeled with [α - 32 P]dCTP (Prime-It II random primer labeling kit; Stratagene, La Jolla, CA) and purified with a Sephadex G-50 column (Sigma-Aldrich, St. Louis, MO) as previously described (23). *T. whipplei* RNA was denatured by 60 min of heating at 50°C before migration into a 1.2% agarose gel. Following 4 to 4.5 h of migration (60 V), RNA was transferred onto a Hybond-N⁺ membrane (Amersham Biosciences, Orsay, France), fixed for 2 h at 80°C, and cross-linked (0.12 J/cm²). Hybridized [α - 32 P]dCTP-labeled fragments were detected by autoradiography (Hyperfilm ECL; Amersham Biosciences).

Sequencing of the 1,500-bp RNA region. Total RNA was separated by polyacrylamide-urea electrophoresis, and the 1,500-base region resolved was excised and resuspended in 200 μ l DEPC-treated H₂O. The RNA solution was dialyzed onto a 0.025- μ m membrane (Millipore) to eliminate urea. One-step RT-PCR was performed using the same primers that were used for Northern blotting in order to amplify total 23S rRNA, 5' and 3' 23S rRNA extremities, and 16S rRNA. The resulting amplicons were sequenced (3100 genetic analyzer) and analyzed using BLASTN (<http://www.ncbi.nlm.nih.gov>).

Phylogenetic analysis. 23S rRNA IVSs were searched in the NCBI databank. Alignment of selected nucleotide sequences was performed with CLUSTAL W (26). A maximum-parsimony analysis was performed using the MEGA2 software (available online). The alignment was sampled for 100 bootstrap replicates.

RESULTS

Extraction of *T. whipplei* RNA. In order to optimize RNA extraction, several volumes (range, 1 to 300 ml) of mid-log cultures of *T. whipplei* strain Twist (12), grown under axenic conditions (22), were treated. Various lysis methods coupled with four distinct extraction procedures, including the Trizol reagent, an RNeasy mini kit (QIAGEN), a FastRNA Pro Blue kit (Qbiogene, Irvine, CA), and a RiboPure bacterial kit (Ambion), were tested. To determine the quality of the extracted RNA, as well as the output of the extraction, all samples were analyzed with a Bioanalyzer 2100. The results obtained clearly demonstrated that extraction performed using the Trizol reagent, starting with 100-ml mid-log cultures of *T. whipplei* sonicated twice for 30 s, was the only method tested that produced detectable quantities of RNA (approximately 5.5 μ g total RNA per 2×10^8 bacteria).

Qualitative analysis of *T. whipplei* RNA. A representative electropherogram pattern for *T. whipplei* RNA obtained by this method is shown Fig. 1A. Three major peaks were observed; the first one (on the left, 24 s) corresponds to the standard calibration. The sizes of the other peaks are approximately 120, 1,500, and 3,100 bp, which correspond to the expected sizes of 5S rRNA, 16S rRNA, and 23S rRNA from *T. whipplei*, respectively (21). Therefore, it is evident that the middle peak, thought to correspond to 16S rRNA, is atypical and composed of two peaks that are close together. Since samples were treated with RNase-free DNase (QIAGEN), the hypothesis that there was contaminating DNA was discarded, which was confirmed by a PCR assay. As shown in Fig. 2, RT-PCR amplification of the *rpoB* gene from our purified RNA was successful, and no amplification was observed in the absence of reverse transcriptase. The atypical electrophoresis profile, cou-

FIG. 1. *T. whipplei* RNA. (A) Representative electropherogram of purified RNA. (B) Schematic representation of *T. whipplei* 23S rRNA (*rrl*). The size of the mature 23S rRNA is 3.1 kb. The position of the insertion sequence (IS) is indicated, as are the positions of the primers used to amplify the probes used for the Northern blot analysis. (C) Agarose gel electrophoresis and Northern hybridization of total RNA from *T. whipplei*. Lanes 1 to 5, Northern hybridization of 32 P-labeled probes complementary to full-length 23S rRNA, 5' end 23S rRNA, 3' end 23S rRNA, 16S rRNA, and 5S rRNA, respectively; lane A, 1.2% agarose gel stained with ethidium bromide.

pled with the fact that in all our experiments the 23S rRNA/16S rRNA ratio, which normally is between 1 and 2 (25), was only 0.19 ± 0.01 (mean $\pm$ standard deviation; $n = 20$), suggested that *T. whipplei* 23S rRNA was effectively cleaved.

Northern blot analysis and sequencing of *T. whipplei* RNA. To check for possible IVS excision from 23S rRNA, Northern

FIG. 2. RT-PCR amplification of purified *T. whipplei* RNA. Amplification was performed using primers TwrpoB.F and TwrpoB.R targeting a 507-bp fragment of the β -subunit of the RNA polymerase gene (*rpoB*). Lane L, 100-bp DNA ladder (Invitrogen Life Technologies); lane 1, RT-PCR; lane 2, negative control with recombinant *Taq* DNA polymerase (Gibco BRL) and without reverse transcriptase. For details see the Materials and Methods.

blotting was performed using specific probes complementary to the full gene or, alternatively, to the 5' or 3' extremities (Fig. 1B). 16S rRNA (*rrs*) and 5S rRNA (*rrf*) probes were also used as controls. As shown in Fig. 1C, when either 16S rRNA or 5S rRNA probes were used (lanes 4 and 5), there was hybridization to a single band located at the expected size of the full-length molecule (i.e., 1,525 and 120 bp, respectively). In contrast, several fragments were observed with 23S rRNA probes (lanes 1 to 3). A 3,100-bp band whose intensity was low, which likely corresponded to unfragmented *rnl*, was recognized by the three distinct *rnl* probes. This was also the case for a 1,500-bp band. *T. whipplei* IVS excision could generate two distinct fragments (1,540 and 1,480 bp, corresponding to the 5' and 3' *rnl* extremities, respectively). These fragments were probably too close together to be resolved in agarose gels, but their recognition by either the 5' or 3' end *rnl* probes demonstrated that splicing had occurred. This was confirmed by amplification and sequencing of the RNA present in this 1,500-base region. Using polyacrylamide gel-extracted RNA as the template and the same primers that were used for the Northern blot probes, we failed to amplify the full-length 23S rRNA. In contrast, amplicons were obtained using primers designed to amplify both 5' and 3' extremities of 23S rRNA, as well as 16S rRNA (data not shown). Their identities were confirmed by BLAST analysis of corresponding sequences (not shown).

Phylogenetic analysis of 23S rRNA IVSs. We analyzed the phylogenetic relationship of the *T. whipplei* IVS identified here with other IVSs found in the 23S rRNA of other bacteria. Among all the candidates identified, *T. whipplei* is the only gram-positive microorganism. Representative nucleotide sequences of each bacterial species were then aligned to construct the maximum-parsimony tree. The topology of the phylogenetic tree obtained is shown in Fig. 3.

DISCUSSION

In this paper we describe an efficient and reproducible protocol for extracting RNA from *T. whipplei*. The best results

FIG. 3. Phylogenetic relationships of *T. whipplei* IVS and selected IVSs from other bacteria, including *Campylobacter jejuni* (accession no. L33972), *S. enterica* (AF227875), *S. enterica* serovar Typhimurium (U49926), *S. enterica* serovar Typhi (U54698), *Proteus vulgaris* (AF176789), *Haemophilus influenzae* (AF090105), *Yersinia enterocolitica* (M35805), *Coxiella burnetii* (X79704), *Helicobacter mustelae* (Z71985), *Helicobacter muridarum* (Z71984), and *Helicobacter canis* (Z71983). The dendrogram was constructed by comparing IVSs using the neighbor-joining method and the p-distance algorithm. Bootstrap values greater than 80% are indicated at the nodes. Scale bar = 5% nucleotide sequence divergence.

were obtained by sonication of bacteria previously resuspended in Trizol reagent, but the resulting RNA profile was not the expected profile (i.e., 120 bp for 5S rRNA, 1,525 bp for 16S rRNA, and 3,102 bp for 23S rRNA) (21). The electropherogram pattern was atypical, suggesting that there was possible RNA cleavage. By using Northern blot analysis and sequencing of the 1,500-base RNA region resolved by electrophoresis, we demonstrated that excision of an IVS postulated to be 80 bp long (10) occurred, leading to cleavage of the 23S rRNA. The sporadic distribution of IVSs throughout the bacteria suggests that their formation results from lateral transfer events (2, 16, 18). We thus examined the phylogenetic relationships between *T. whipplei* and other characterized bacterial IVSs, which confirmed that *T. whipplei* 23S rRNA IVS was acquired by lateral gene transfer between enteric bacteria.

The function of IVSs is poorly documented. It has been demonstrated that an *S. enterica* serovar Typhimurium RNase III (*mc*)-deficient strain in which IVSs are not excised grows slowly. Nevertheless, this effect was not attributed to a fault in IVS excision but rather to the absence of other RNase III-catalyzed reactions (15). This is consistent with the fact that when *Escherichia coli* (which does not normally contain IVSs) is transformed with a plasmid carrying an IVS-containing *S. enterica* serovar Typhimurium *rnl* gene, the 23S rRNA is fragmented but the growth rate is similar to that of the wild type (9). It has also been reported that, in contrast to what was observed in *E. coli*, 23S rRNA of *Salmonella* strains is rapidly degraded when growing cells enter the stationary phase. This process, which in turn regulates rRNA and ribosome synthesis, might facilitate adaptation of the microorganisms to a rapidly changing growth environment (11). Such adaptive pathways should be useful for *T. whipplei*, which it is thought to have an environmental reservoir (13). Recent work highlighted the finding that IVSs are mainly localized in symbionts and pathogens of eukaryotic hosts (2). From this observation it was hypothesized that IVSs could result from close bacterium-host relationships that lead to genome reduction as a part of an

overall adaptation process. The pronounced secondary structure of IVSs may promote increased communication between bacteria and their host cells (2). A functional role in a pathway independent of the ribosome but transcriptionally linked to protein production in the bacterium and also in the host has thus been evoked (2). While *T. whipplei* can be cultivated under axenic conditions in vitro (22), this microorganism is an intracellular pathogen (12) with a genome size of only 928 kb (3, 21). This bacterium is deficient in essential amino acid pathways (22), which are provided by the host, like other small-genome intracellular bacteria. It could be hypothesized that excision of the IVS from 23S rRNA should enhance transcription of these nutrients from the host cell, thus ensuring the survival of *T. whipplei*. Although the precise function of IVSs remains to be elucidated (18), we believe that it is important to know which sequences are removed from the primary rRNA transcript during rRNA maturation since rRNA sequences are widely used for bacterial phylogeny and identification purposes.

ACKNOWLEDGMENTS

We greatly appreciate the assistance of Guy Longepied for technical assistance with the Northern blot assay. We also thank Michael Mitchell for revising the English of the manuscript.

N.C. received financial support from the Programme Hospitalier de Recherche Clinique (PHRC) from the French Ministry of Health.

REFERENCES

- Andersson, S. G., A. Zomorodipour, J. O. Andersson, T. Sicheritz-Ponten, U. C. Alsmark, R. M. Podowski, A. K. Naslund, A. S. Eriksson, H. H. Winkler, and C. G. Kurland. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**:133–140.
- Baker, B. J., P. Hugenholtz, S. C. Dawson, and J. F. Banfield. 2003. Extremely acidophilic protists from acid mine drainage host *Rickettsiales*-lineage endosymbionts that have intervening sequences in their 16S rRNA genes. *Appl. Environ. Microbiol.* **69**:5512–5518.
- Bentley, S. D., M. Maiwald, L. D. Murphy, M. J. Pallen, C. A. Yeats, L. G. Dover, H. T. Norbertczak, G. S. Besra, M. A. Quail, D. E. Harris, A. von Herbay, A. Goble, S. Rutter, R. Squares, S. Squares, B. G. Barrell, J. Parkhill, and D. A. Relman. 2003. Sequencing and analysis of the genome of the Whipple's disease bacterium *Tropheryma whipplei*. *Lancet* **361**:637–644.
- Burgin, A. B., K. Parodos, D. J. Lane, and N. R. Pace. 1990. The excision of intervening sequences from *Salmonella* 23S ribosomal RNA. *Cell* **60**:405–414.
- Drancourt, M., A. Carlioz, and D. Raoult. 2001. *rpoB* sequence analysis of cultured *Tropheryma whippelii*. *J. Clin. Microbiol.* **39**:2425–2430.
- Evguenieva-Hackenberg, E., and G. Klug. 2000. RNase III processing of intervening sequences found in helix 9 of 23S rRNA in the alpha subclass of *Proteobacteria*. *J. Bacteriol.* **182**:4719–4729.
- Fenollar, F., M. L. Birg, V. Gauduchon, and D. Raoult. 2003. Culture of *Tropheryma whipplei* from human samples: a 3-year experience (1999 to 2002). *J. Clin. Microbiol.* **41**:3816–3822.
- Fenollar, F., P. E. Fournier, D. Raoult, R. Gerolami, H. Lepidi, and C. Poyart. 2002. Quantitative detection of *Tropheryma whipplei* DNA by real-time PCR. *J. Clin. Microbiol.* **40**:1119–1120.
- Gregory, S. T., M. O'Connor, and A. E. Dahlberg. 1996. Functional *Escherichia coli* 23S rRNAs containing processed and unprocessed intervening sequences from *Salmonella typhimurium*. *Nucleic Acids Res.* **24**:4918–4923.
- Hinrikson, H. P., F. Dutly, and M. Altwegg. 2000. Analysis of the actinobacterial insertion in domain III of the 23S rRNA gene of uncultured variants of the bacterium associated with Whipple's disease using broad-range and "*Tropheryma whippelii*"-specific PCR. *Int. J. Syst. Evol. Microbiol.* **50**:1007–1011.
- Hsu, D., L. M. Shih, and Y. C. Zee. 1994. Degradation of rRNA in *Salmonella* strains: a novel mechanism to regulate the concentrations of rRNA and ribosomes. *J. Bacteriol.* **176**:4761–4765.
- La Scola, B., F. Fenollar, P. E. Fournier, M. Altwegg, M. N. Mallet, and D. Raoult. 2001. Description of *Tropheryma whipplei* gen. nov., sp. nov., the Whipple's disease bacillus. *Int. J. Syst. Evol. Microbiol.* **51**:1471–1479.
- Maiwald, M., F. Schuhmacher, H. J. Ditton, and A. von Herbay. 1998. Environmental occurrence of the Whipple's disease bacterium (*Tropheryma whippelii*). *Appl. Environ. Microbiol.* **64**:760–762.
- Marth, T., and D. Raoult. 2003. Whipple's disease. *Lancet* **361**:239–246.
- Mattatall, N. R., and K. E. Sanderson. 1998. RNase III deficient *Salmonella typhimurium* LT2 contains intervening sequences (IVSs) in its 23S rRNA. *FEMS Microbiol. Lett.* **159**:179–185.
- Miller, W. L., K. Pabbaraju, and K. E. Sanderson. 2000. Fragmentation of 23S rRNA in strains of *Proteus* and *Providencia* results from intervening sequences in the *rrn* (rRNA) genes. *J. Bacteriol.* **182**:1109–1117.
- Palmer, G. H. 2002. The highest priority: what microbial genomes are telling us about immunity. *Vet. Immunol. Immunopathol.* **85**:1–8.
- Pronk, L. M., and K. E. Sanderson. 2001. Intervening sequences in *rrl* genes and fragmentation of 23S rRNA in genera of the family *Enterobacteriaceae*. *J. Bacteriol.* **183**:5782–5787.
- Ralph, D., and M. McClelland. 1993. Intervening sequence with conserved open reading frame in eubacterial 23S rRNA genes. *Proc. Natl. Acad. Sci. USA* **90**:6864–6868.
- Raoult, D., M. L. Birg, B. La Scola, P. E. Fournier, M. Enea, H. Lepidi, V. Roux, J. C. Piette, F. Vandenesch, D. Vital-Durand, and T. J. Marrie. 2000. Cultivation of the bacillus of Whipple's disease. *N. Engl. J. Med.* **342**:620–625.
- Raoult, D., H. Ogata, S. Audic, C. Robert, K. Suhre, M. Drancourt, and J. M. Claverie. 2003. *Tropheryma whipplei* Twist: a human pathogenic actinobacteria with a reduced genome. *Genome Res.* **13**:1800–1809.
- Renesto, P., N. Crapoulet, H. Ogata, B. La Scola, G. Vestris, J. M. Claverie, and D. Raoult. 2003. Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. *Lancet* **362**:447–449.
- Renesto, P., P. Dehoux, E. Guin, L. Touqui, P. Cossart, and D. Raoult. 2003. Identification and characterization of a phospholipase D-superfamily gene in rickettsiae. *J. Infect. Dis.* **188**:1276–1283.
- Rovero, C., P. Renesto, N. Crapoulet, K. Matsumoto, P. Parola, H. Ogata, and D. Raoult. 2005. Transcriptional response of *Rickettsia conorii* exposed to temperature variation and stress starvation. *Res. Microbiol.* **156**:211–218.
- Skrypina, N. A., A. V. Timofeeva, G. L. Khaspekova, L. P. Savochkina, and R. S. Beabealashvili. 2003. Total RNA suitable for molecular biology analysis. *J. Biotechnol.* **105**:1–9.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 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.