We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Molecular Phylogenetic Identification of Actinobacteria

Xiu Chen, Yi Jiang, Qinyuan Li, Li Han and Chenglin Jiang

Additional information is available at the end of the chapter

Abstract

Molecular phylogenetics plays an important role in prokaryote taxonomy and identification. The content of this chapter is to introduce the common application of genetic criteria including 16S rRNA gene sequence nucleotide similarity and phylogeny, DNA G+C content, and DNA–DNA hybridization. However, the genomics era might put forward some new criteria. This chapter emphasizes the methods and basic principles of molecular identification and taxonomy of actinobacteria.

Keywords: 16S rDNA, molecular phylogenetic, genetic criteria

1. Introduction

Currently, the taxonomy and identification of prokaryotes rely on polyphasic combinations of phenotypic, chemotaxonomic and genotypic characteristics. Initially, taxon of actinobacteria is based on phenotypic markers such as morphology, growth requirements or pathogenic potential [1]. Later, physiological and biochemical properties of bacteria were also used for this purpose [2, 3]. Chemotaxonomy [4] and DNA–DNA hybridization techniques [5, 6] were widely used subsequently. The advent of DNA amplification and sequencing techniques, in particular of the 16S rRNA gene, constituted the crucial criteria forward for determining the taxonomic status of prokaryotes [7–9], greatly increased the rate of discovering novel species [10] and now routinely carried out as the first step in identifying novel organisms [11–13]. 16S rRNA gene was the best target molecule for studying the phylogenetic relationships because it is present in all the bacteria, functionally constant and composed of highly conserved as well as more variable regions. Some other molecular methods have been used in the classification of prokaryotes, such as multilocus sequencing typing (MLST) [14,15], SDS-PAGE analysis of whole cell soluble proteins [16], secondary structure and signature nucleotides analysis of variable areas of the 16S rRNA gene [17,18]. However, genomic age put forward that some

genomic characteristics have great potential in the taxonomy of bacteria and archaea as a substitute for the traditional method of determination of G+C content mol% and the labour-intensive DNA–DNA hybridization (DDH) technique [19-22].

2. Extraction and purification of genomic DNA

DNA is the carrier of genetic information and also the basis of gene expression. Molecular phylogenetic is the basic method for identification of actinobacteria. Before genetic-based methods, the first thing is the extraction and purification of DNA, and the quality of DNA is the premises to the success or failure of the experiment.

2.1. The principle of extraction and purification of genomic DNA

DNA contains all the genetic information which is all stored in the primary structure of DNA. Therefore, to ensure the quality of the DNA in the preparation of DNA samples is of great significance. Otherwise, it is difficult to get the right result. To ensure the quality of DNA, the following should be noticed: firstly, to avoid high temperature; secondly, to control the pH at a certain pH range (pH 5–9); Thirdly, maintain the ionic strength of buffer which is of significance to maintain the space configurations of DNA; And lastly, reduce the disruption of DNA in the course of extraction by physical factors, such as high speed oscillation, mixing and freezing–thawing. There are a lot of DNA enzymes in the environment that can digest the DNA or RNA, therefore some material used in the extraction has to be sterilized and enzyme inhibitors should be added in the extraction buffer at the same time. In addition, avoiding contamination of exogenous DNA is also important.

2.2. The main steps of extraction and purification of genomic DNA

2.2.1. Cell disruption

Genomic DNA is an intracellular constituent, so the first step of genomic DNA extraction is cell disruption. For cell disruption of microbial cells, the following several kinds of methods are commonly used: enzyme digestion, ultrasonic, grinding with liquid nitrogen, alkali treatment, microwave preparation, freeze–thawing and surfactant treatment.

2.2.2. Removal of nucleoprotein from genomic DNA

The binding force of nucleic acid and protein is mainly electrostatic forces, hydrogen bonding and Van der Waals interactions. The most difficult thing in the extraction is to separate the closely integrated protein from genomic DNA and avoid the degradation of DNA. There are some commonly used methods such as adding the concentrated solution of NaCl, which makes the nucleoprotein to depolymerize. Adding the SDS makes the protein free from genomic DNA in phenol/chloroform extraction. However, there are many kits which can remove nucleoprotein from the genomic DNA, and the result is always optimized.

2.2.3. Precipitation of genomic DNA

Precipitation is the best way to concentrate DNA and is widely used. The advantage of precipitation is that it can remove some salt ions from the solution. It is also a step for nucleic acid purification. Ethanol, isopropanol and polyethyleneglycol (PEG) are commonly used for DNA precipitation. Ethanol is the most preferred precipitant. Two times the volume of ethanol is effective for precipitation of DNA and 2.5 times for RNA if under the appropriate salt concentration. The advantage of isopropanol is a small volume requirement and it is suitable for large DNA samples in a low concentration. The disadvantage of isopropanol is that it is easier to make salt coprecipitation with DNA and is difficult to volatilize. So washing with 70% ethanol several times to remove the isopropanol and salt is necessary. The PEG can select DNA fragments of different length. In addition, $MgCl_2$, NaAC, KAC, NH_4AC, Nil and LiCl are useful as assisted components.

2.2.4. Time and temperature of the nucleic acid precipitation

It is generally believed that the nucleic acid precipitation should be carried out at low temperature, such as –20℃ or –70℃ for a few hours or even an overnight. But this kind of treatment is easy to cause the salt coprecipitation with DNA, so at 0℃ or 4℃ for 30–60 min is recommended.

2.3. Some specific methods for extraction and purification of genomic DNA

2.3.1. Enzymatic disruption method

This method is frequently used and suitable for most actinobacteria.

i.   50 mg of the pretreated cell samples are suspended in 480 µl TE buffer and 20 µl of lysozyme solution (50 mg/ml), then put them in shaker at 37℃ for overnight.

ii.   Add 50 µl of 20% SDS and 5 µl of proteinase K (20 mg/ml), vortexing for 1 min, then put incubate at 55℃ for 60 min.

iii.   Supernatants are transferred to fresh microcentrifuge tubes after centrifugation at 10,000 × g for 5 min at room temperature.

iv.   Add 550 µl of mixer of phenol: chloroform: isoamyl alcohol (25:24:1), vortexing for 1 min, centrifuge for 10 min (12,000 rpm). The aqueous phase containing the community DNA was transferred to another sterile microcentrifuge tube.

v.   Repeat IV.

vi.   Add 50 µl of 3 mol/L sodium acetate (pH 4.8–6.2), vortex gently, add 500 µl of isopropanol, or 800 µl ethylalcohol, then keep in the room temperature for more than 10 min or put it at 4℃ for 30 min to 2 h or overnight if necessary.

vii.   Centrifuge for 10 min (12,000 rpm), discard the supernatant, add 200 µl of 70% ethanol, shake slightly, then centrifuge for 5 min (12,000 rpm), discard the ethanol. Repeat 1–2 times. Then dry at room temperature or at a higher temperature (≤55°C).

viii. Add 50 µl of 1×TE buffer to dilute the DNA (depend on the volume of DNA), preserve it at –20℃ for later use.

2.3.2. Extraction of genomic DNA using chelex-100

This method is fast, simple and convenient, but the extracted DNA is not suitable for long periods storing.

i. 5–10 mg of the pretreated cell samples are suspended with 50 µl chelex buffer (5% Chelex-100), incubate at 100°C for 20–40 min.

ii. Centrifuge for 10 min (12,000 rpm), then the supernatant was transferred to another sterile microcentrifuge tube, and keep it at 4℃ or –20℃ for later use.

2.3.3. Extraction of genomic DNA using microwave

i. 50 mg of the pretreated samples are suspended in 1 ml washing buffer.

ii. Centrifuge (5,700 rpm) for 1 min, discard the supernatant.

iii. Add 50 µl lysis buffer, vortex for 30 s, then treat with microwave at 600 w for 45 s.

iv. Add 500 µl preheat (65℃) extraction buffer, vortex for 5 s.

v. Follow step IV–VIII in Section 6.2.3.1

2.3.4. Extraction of genomic DNA by grinding with liquid nitrogen

This method is always used for mass extraction of DNA.

i. Put 1–2 g of wet cell mass into mortar, take a suitable amount of liquid nitrogen to cover the mass, and grind to freeze.

ii. Repeat grind for 4–5 times.

iii. Transfer the product to a sterile microcentrifuge tube (50 ml) with 7 ml TE buffer.

iv. Add 700 µl 20% SDS and 800 µl of proteinase K (20 mg/ml), vortex for 1 min, then incubate at 55℃ for 60 min (the final concentration of proteinase K is 20 µg/ml).

v. Supernatants are transferred to another tube after a centrifugation at 10,000×g for 5 min at room temperature (optionally).

vi. Add 8 ml of mixer of phenol: chloroform: isoamyl alcohol (25:24:1), vortex for 2 min, centrifuge for 10 min (12,000 rpm); the aqueous phase containing DNA is transferred to another sterile microcentrifuge tube (50 ml) (do not suck the waste in the middle).

vii. Repeat IV.

viii. Add 800 µl of 3 mol/L sodium acetate (pH 4.8–6.2) into the supernatant, vortex gently, then add 8 ml of isopropanol, or 16 ml of absolute ethyl alcohol, keep in the room temperature for more than 10 min or put it at 4℃ for 30 min to 2 h or overnight if necessary.

ix. Centrifuge for 10 min (12,000 rpm), discard the supernatant, add 4 ml of 70% ethanol, shake slightly, centrifuge it for 5 min (12,000 rpm), then discard the ethanol. Repeat 1–2 times, dry at room temperature or at a higher temperature (≤55°C).

x. Add 1×TE buffer (≥1 ml) to dilute the DNA (depend on the volume of DNA), preserve it at –20℃ for later use.

2.4. Purification of genomic DNA

A high purity of DNA is necessary for determination of G+C content, DNA–DNA hybridization and sequencing. The following protocol could be a reference.

i.   Add 480 μl 1×TE buffer to resuspend the extracted DNA from 100 mg of cell mass.

ii.   Add 5 μl of proteinase K (20 mg/ml) and 15 μl RNase A (400 μg/ml), incubate at 37℃ for 30–60 min.

iii.   Add 550 μl of mixer of phenol: chloroform: isoamyl alcohol (25:24:1), vortex for 2 min, centrifuge for 10 min (12,000 rpm), the aqueous phase containing DNA was transferred to another sterile microcentrifuge tube after centrifugation (do not suck the waste in the middle).

iv.   Add 550 μl chloroform, vortex for 2 min, centrifuge for 10 min (12,000 rpm), and the aqueous phase containing DNA is then transferred to another sterile microtube.

v.   Add 50 μl of 3 mol/L sodium acetate (pH 4.8–6.2) into the supernatant, vortex gently and then add 500 μl of isopropanol, or 800 μl absolute ethyl alcohol, vortex gently again, keep at room temperature for more than 10 min or put it at 4℃ for 30 min to 2 h or overnight if possible.

vi.   Centrifuge for 10 min (12,000 rpm), discard the supernatant, add 200 μl of 70% ethanol, vortex slightly, centrifuge it for 5 min (12,000 rpm), discard the ethanol. Repeat 2–3 times. dry at room temperature or at a higher temperature (≤55°C).

vii.   Add 50 μl of 0.1×SSC or deionized water to dilute the DNA (depend on the volume of DNA), preserve it at –20℃ for later use.

3. Amplification of 16S rDNA sequence

Polymerase chain reaction (PCR) is an ingenious technique used to exponentially amplify a specific target DNA sequence. PCR was developed by Kary Mullis in 1983. He won a Nobel Prize in chemistry in 1993 for his invention. PCR has been elaborated in many ways since its introduction and is now commonly used for a wide variety of applications including genotyping, cloning, mutation detection, sequencing, microarray, forensics and paternity testing.

Typical PCR is a three-step reaction (Figure 1.). The sample containing a dilute concentration of template DNA is mixed with a heat-stable DNA polymerase, primers, deoxynucleoside triphosphates (dNTPs) and buffer (including magnesium). In the first step of PCR, the sample is heated at 94–98℃ for 3–8 min, which pre-denatures the double-stranded DNA and splitting it into two single strands. In the second circulate step, the sample is heated at 94–98℃ for 30–60 s to denature the double-stranded DNA continually, then the temperature is decreased to approximately 52–65℃ (depend on the annealing temperature of primers) to allow the primers to bind or anneal with specific site in single strands which is also known as the template. Lastly, the temperature is typically increased to 72℃, allowing the DNA polymerase to react by the addition of dNTPs to create a new strand of DNA. The times of extension are varied depending upon the length of target sequence and the kind of polymerase. Generally, the extend speed

of *Taq*-polymerase is 1 kb/min. In the third step, it is a final extension, which is to repair and to fill some gaps of the products in the second step, and the reaction rate reaches a plateau in this step.
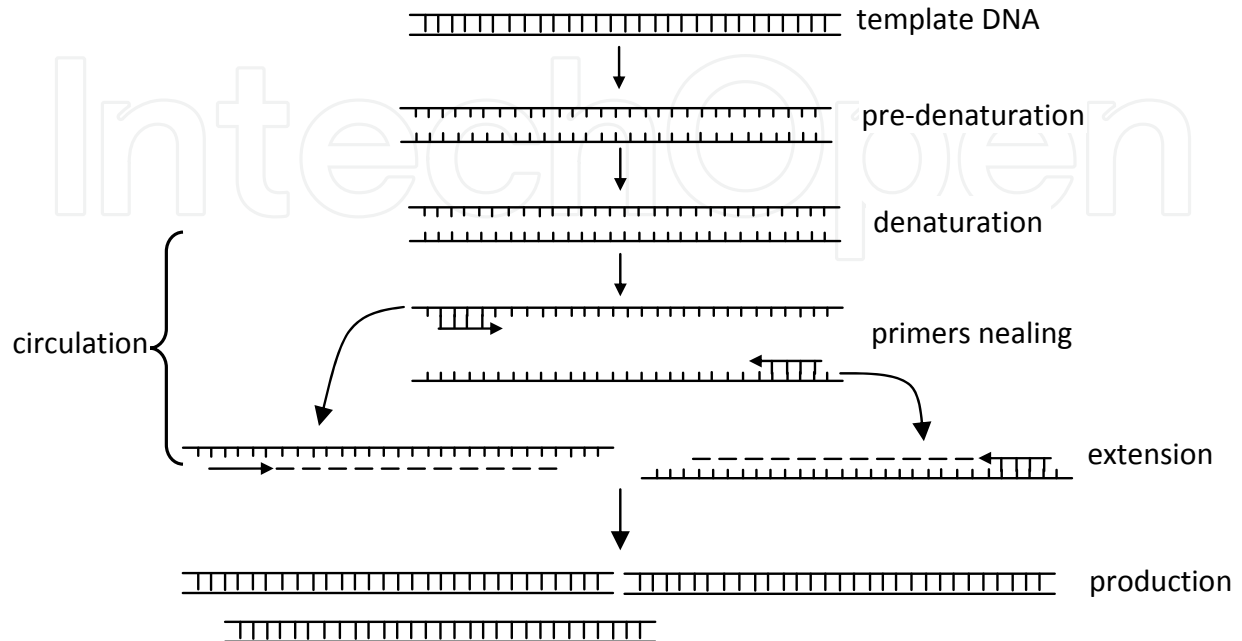

Figure 1. The principle of DNA amplification

3.1. Amplification of 16S rDNA

Universal primers of 16S rDNA for actinobacteria:

27F: AGAGTTTGATCCTGGCTCAG

1492R: GGTTACCTTGTTACGACTT

16S rRNA gene is the best target for studying the phylogenetic relationships because it is present in all bacteria, functionally constant, composed of highly conserved as well as more variable regions. As described above, determination of 16S rDNA sequence is routinely carried out as the first step in identifying novel organisms. The ingredients for amplification of 16S rDNA are listed in Table 1..

Generally, the condition for 16S rDNA amplification is:

1.   94℃ 4 min

2.   94℃ 45 s

3.   55℃ 45 s

4.   72℃ 90 s

5.   Repeat steps 2–4 for 35 times

Ingredient*	Addition
10×PCR buffer	5.0 μl
dNTPs	4.0 μl
27F (25 pmol/μl)	1.0 μl
1492R (25 pmol/μl)	1.0 μl
Template DNA	1.0 μl
Taq DNA polymerase	0.3 μl
dd H$_2$O	37.7 μl
Total volume	50 μl

*The ingredients of the system are bought from TaKaRa.

Table 1. Composition and dosage of amplification

6. 72℃ 10 min

7. 4℃ hold (optional)

3.2. Potential problems in amplification of 16S rDNA sequence

i.   Positive and negative controls must be used and run every time; if the negative controls become positive, the amplification should be carried out again.

ii.   No products or all the stripes are weak. Higher temperatures and long reaction time for high GC content should be optional. Increase the dose of polymerase and template DNA and the number of cycles. Reducing the annealing temperature might be also useful. However, checking whether the system is out of date is also necessary.

iii.   Nonspecific products appeared. In contrast to no products, reduce the dose of polymerase and template DNA and raise the annealing temperature.

iv.   Impurities such as phenol or too much salt will result in no PCR product or nonspecific products.

v.   Impure template DNA would result in sequencing failure or double peak for pure strain identification.

vi.   If PCR product of wrong band, likely causes include: incorrect primer, template mutations, contamination and incorrect annealing temperature.

vii.   Too many primers will lead to the annealing of themselves.

3.3. Detection of polymerase amplification products

Since the world's oldest electrophoresis experiment was carried out for nearly 200 years, electrophoresis technology has been continuously improved and developed. Now, electrophoresis is one of the most commonly used methods for biological macromolecule detection

and has played a huge boost. Electrophoresis is a technique also used to purify macromolecules, especially proteins and nucleic acids, which are different in size, charge or conformation. When charged molecules are placed in an electric field, they migrate towards either the positive or negative pole according to their charge. Nucleic acids have a consistent negative charge imparted by their phosphate backbone and migrate towards the anode. Nucleic acids are electrophoresed within a matrix or 'gel'. Commonly, the gel is cast in the shape of a thin slab with wells for loading the sample. The gel is immersed within an electrophoresis buffer (TAE or TBE) that provides ions to carry a current and to maintain the pH at a relatively constant value. The gel itself is composed of either agarose or polyacrylamide, each of which has attributes suitable to particular tasks: agarose is a polysaccharide extracted from seaweed. It is typically used at concentrations of 0.5–2%. Agarose gels are extremely easy to prepare: simply mix agarose powder with buffer solution, melt it by heating and pour the gel. It is also non-toxic. Agarose gels have a large range of separation but with relatively low resolving power. By varying the concentration of agarose, fragments of DNA from about 100 bp to 50,000 bp can be separated using standard electrophoretic techniques. Polyacrylamide is a cross-linked polymer of acrylamide. The length of the polymer chains is dictated by the concentration of acrylamide used, which is typically between 3.5% and 20%. Polyacrylamide gels are significantly more annoying to prepare than agarose gels and have a rather small range of separation, but with very high resolving power. Because oxygen inhibits the polymerization process, they must be poured between glass plates (or cylinders). Acrylamide is a potent neurotoxin and should be handled with care. Wear disposable gloves when handling solutions of acrylamide, and a mask when weighing out powder. In the case of DNA, polyacrylamide is used for separating fragments of less than 500 bp. However, under appropriate conditions, fragments of DNA differing in length by a single base pair are easily resolved. In contrast to agarose, polyacrylamide gels are used extensively for separating and characterizing mixtures of proteins. The protocol of detection of 16S rDNA sequences by agarose gel electrophoresis is:

i.   To slot the organic glass mold in a horizontal position, put the comb in the right position based on your needs.

ii.   Prepare 1.0% (w/v) agarose gels with TAE or TBE buffer and heated by microwave oven.

iii. Add the nucleic acid dye (GoodViewTM, EB-Ethidium bromide, GeneFinder™, or SYBER greenI) to the agarose gels after it cools down (<50℃), mix it gently.

iv.   Pool the mixed agarose gels to the mold; if there are air bubbles, get rid of it.

v.   Pull out the comb slightly after the agarose gels are hardened, make sure the pore is intact.

vi.   Mix 5 μl amplification products with DNA loading buffer (depending on the concentration; 1–2μl for 5×loading buffer) and pipe the mix to the gel pore gently. Add the marker lastly.

vii.   Put the agarose gels which with samples into electrophoresis bath (TAE or TBE) gently, make sure the gel pore is closed to negative pole.

viii.   Run the electrophoresis for about 30 min of a voltage at 4~7 V/cm.

ix.   Take out of the agarose gels and the results were analyzed by gel imaging system.

x.   Send the positive products for sequencing.

3.4. Analysis of 16S rDNA sequence

There are two main cases of 16S rDNA sequence analysis. One is a partial sequence of 16S rDNA sequenced from one direction and there is no need to assemble. Another kind is a contig assembled by two sequences which is always produced by clone to get an almost complete 16S rDNA sequence with high quality. Two types of files will be received from sequencing company, one is ablformat and could be opened by using Chromas, and another is Editseq format which could be opened using Editseq, Bioedit or Notepad. Quality map of sequence is shown in first one and an editable sequence is listed in the later. The qualified sequence is aligned in database (http://www.ezbiocloud.net/eztaxon and http://blast.ncbi.nlm.nih.gov are usually used). The analysis of alignment as well as construction of a phylogenetic tree will be detailed later. The SeqMan in the DNAStar package, Sequencher or vector NTI can be used for assembling. The contig can be assembled by SeqMan as in the following steps:

i.   Open SeqMan, click 'sequence' and then click 'add', add the two sequences of abl format, click 'done' (Figure 2.).
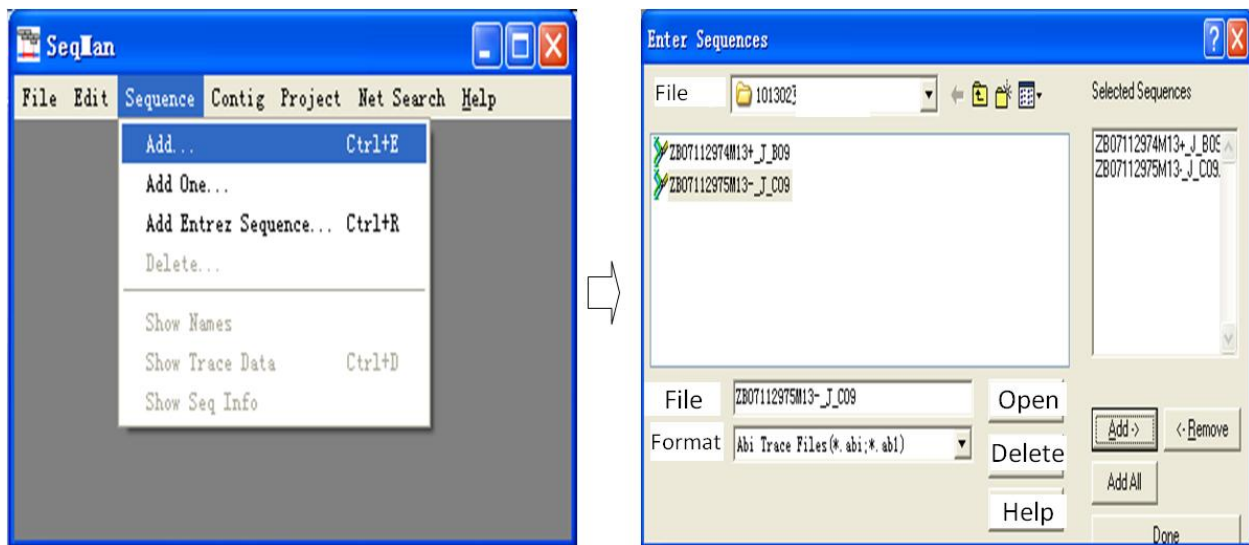
Figure 2. Add sequence to SeqMan

ii.   Click 'assemble', double click the assembled file name to open the contig.
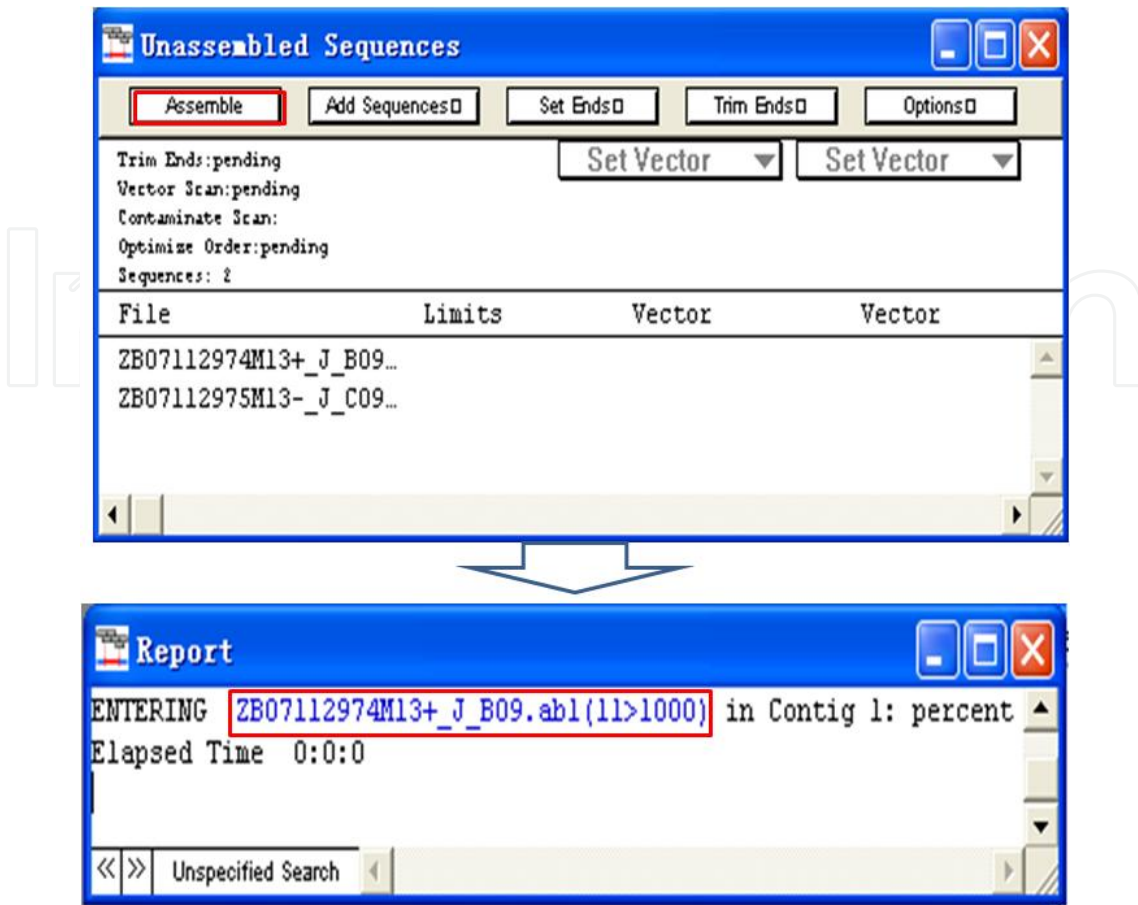
Figure 3. Assemble sequence

iii. Click the '▼' in front of file name to see the quality map; it needs to compare the quality of two maps and to decide which base could be used if the consensus does not match perfect.
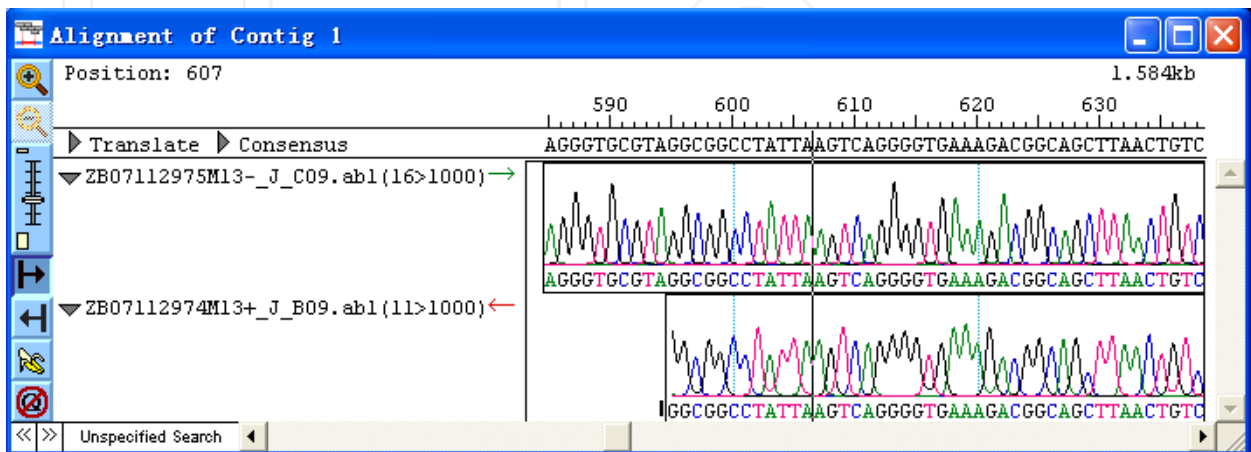
Figure 4. Check the assembled sequence

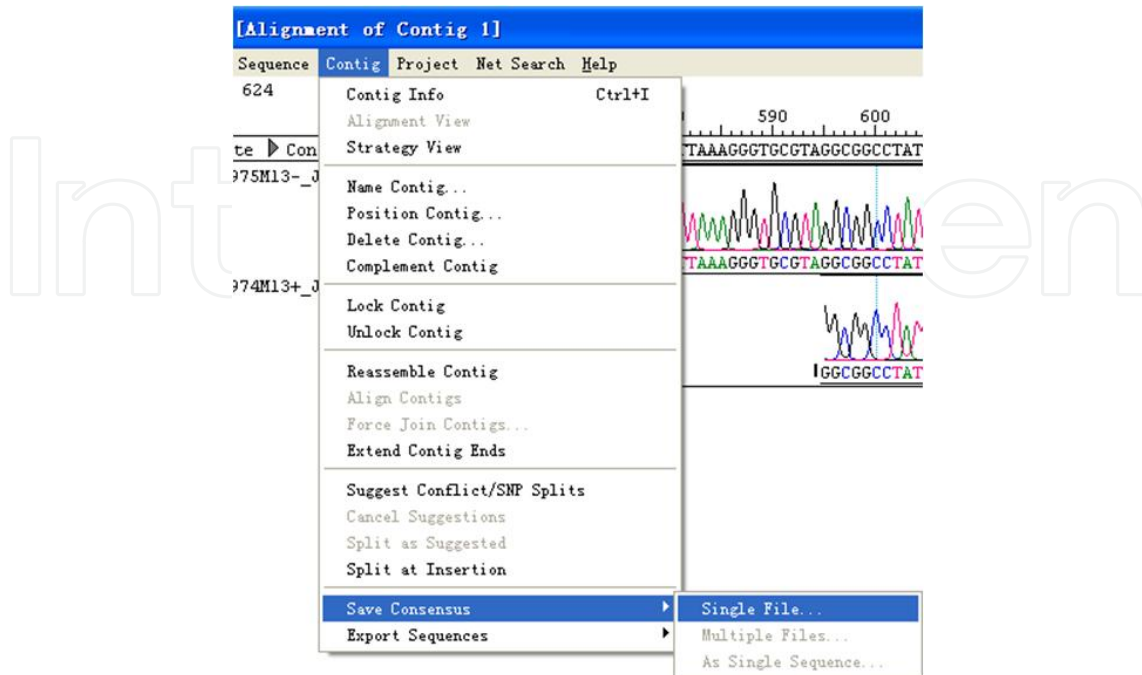**iv.**   Click 'contig', then 'save consensus' and 'single file' to save the result.

Figure 5. Save the assembled sequence

If the contig comes from clone, the vector should be cut out as follows:

i.   Open the webpage (http://www.ncbi.nlm.nih.gov/tools/vecscreen/) and paste
sequence to the following window, click 'Run VecScreen'.

Figure 6. Run vecscreen

ii.   The following is the graphic summary in the report. Cut out the matched sequence, i.e. just the sequence from 37–1,584 could be used to construct a phylogenetic tree.

Figure 7. The report of vecscreen

4. Construction of phylogenetic tree based on 16s rDNA sequences

During the course of evolution, the genes, the numbers of genes, their functions and the sizes of the genomes are continually modified. If genes originate from a common ancestor gene and fulfill the same function in a cell, they are said to be homologous. The degree of divergence between homologous genes is considered a measure for their relatedness. In molecular phylogeny, the relationships among organisms, usually extant, are examined by comparing homologous DNA or protein sequences. The relationships are displayed as phylogenetic trees with branch (or edge) lengths reflecting the degrees of genetic divergence. Each branch tip represents an extant sequence, the internal nodes or vertices represent unknown ancestors to the terminal nodes. The branching pattern and branch lengths describe the evolutionary pathways leading to the sequences at the terminal nodes. Clusters of terminal branches connected to a common ancestor are termed clades [23].

4.1. Access to reference sequences

After alignment in the database (http://www.ezbiocloud.net/eztaxon or http://blast.ncbi.nlm.nih.gov are usually used), the closed bacteria are listed in a column and the sequences of these bacteria can be downloaded.

Access of reference sequences from Ezbiocloud is according to the following steps:

i.   Upload or paste sequence in the place A or C according to the requirement respectively, or type in the accession number of Genbank in place B, click 'identify' to blast the sequence.

Figure 8. Add sequence to Ezbiocloud

ii.   Click query number to view details.

iii.   Click 'FASTA(zZ)' to download the reference sequences file in fasta format.

Figure 9. Download reference sequences from Ezbiocloud

Get reference sequences from NCBI according to the following steps:

i.   Choose 'blastn' in the blast webpage, paste sequence in place A or upload a file in place B, choose others (nr, etc.) in the column of database, then click 'blast' in the program selection to blast.
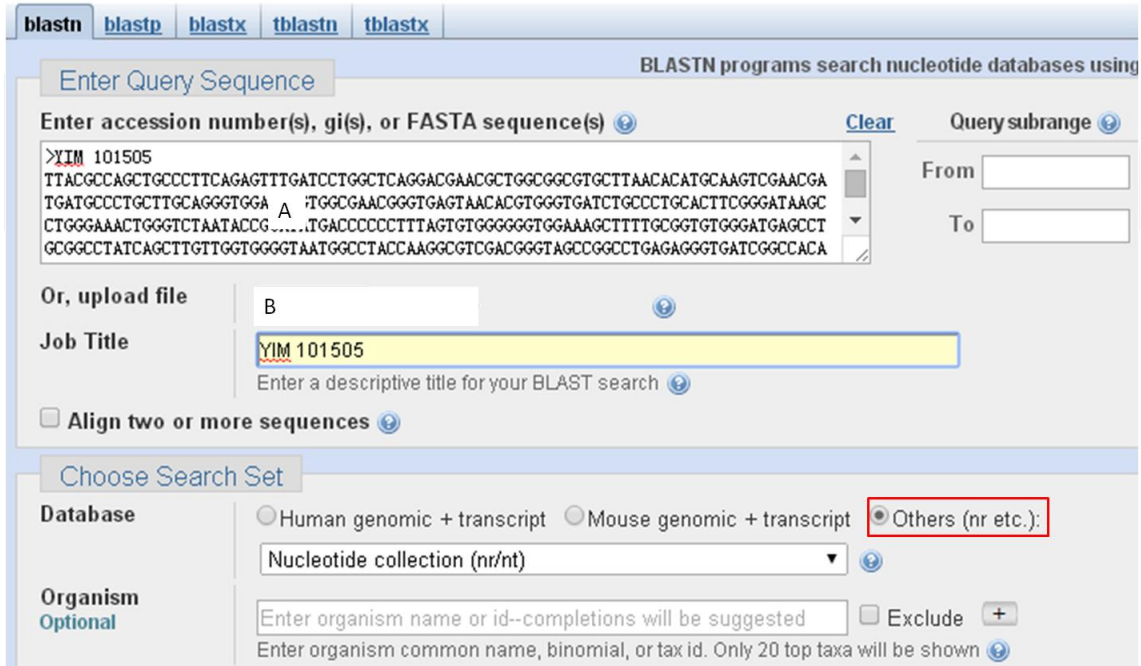
Figure 10. Add sequence to blast of NCBI

ii.   Click the accession number to see details and download sequence one by one, or click the option in the download to download selected sequences.
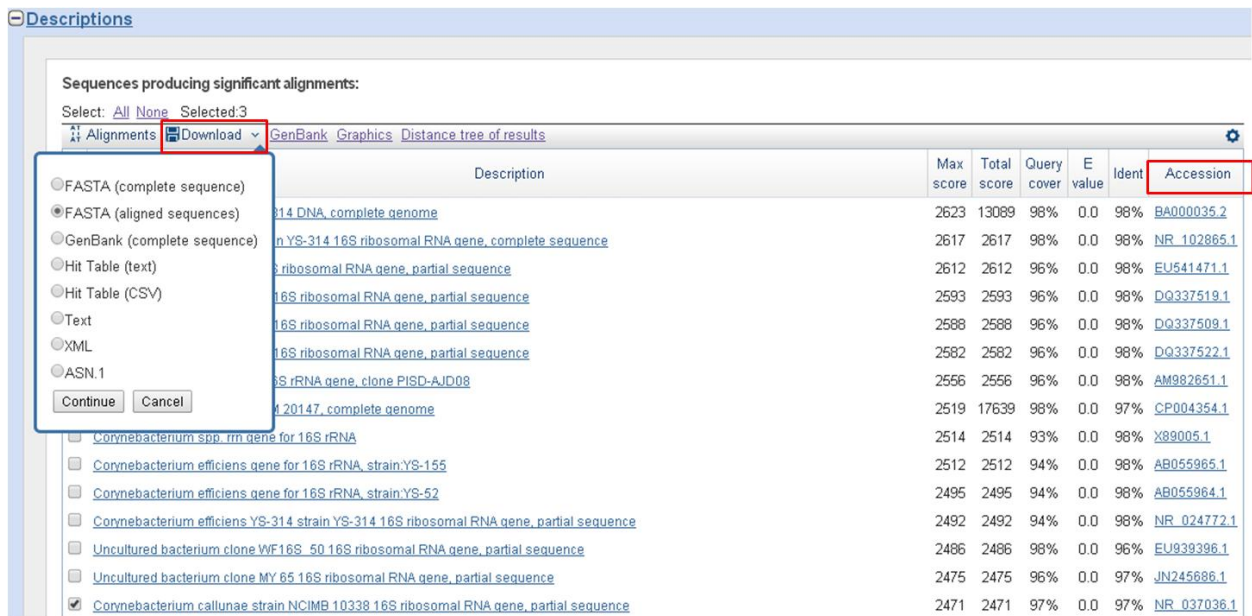
Figure 11. Download reference sequences from NCBI

4.2. Sequences alignment

Prior to the phylogenetic analyses, an alignment of the sequences has to be assembled. If sequences of homologous genes show differences in lengths due to insertions or deletions, gaps have to be inserted to place functionally corresponding positions in the same vertical column of the alignment.

CLUSTAL X and CLUSTAL W are the versions of windows for multiple sequence alignment. CLUSTAL X provides a platform for multiple sequence alignment and analysis results. Users can cut and paste the sequence to change the order, can also realign selected sequences and highlighted low score snippets or abnormal residues, etc. Anyway, the interface of CLUSTAL X is more friendly, intuitive and easier to operate than CLUSTAL W. The basic approach to working in data of CLUSTAL X and CLUSTAL W will be introduced in this part.

Sequences alignment by CLUSTAL-X1.83:

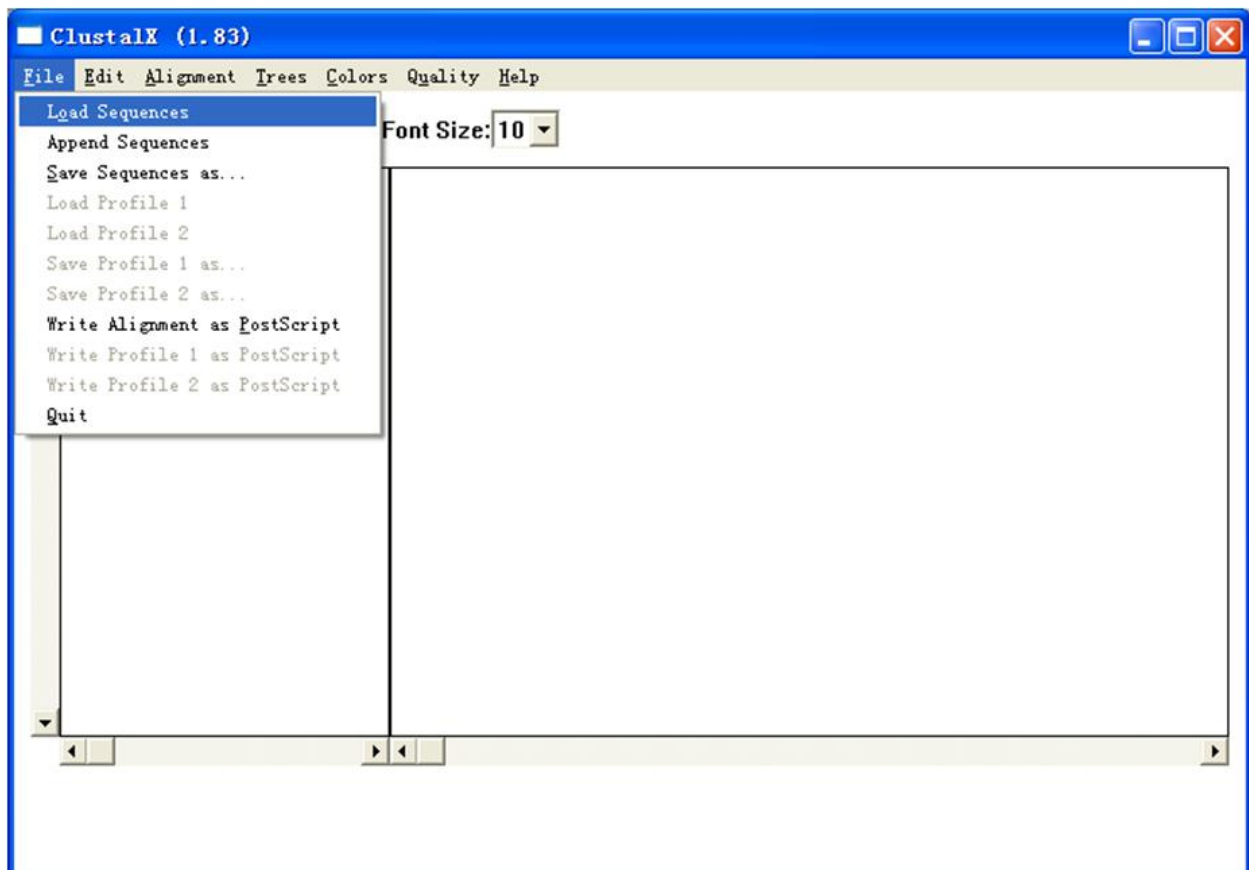**i.**   Load sequences are in fasta, aln or clustal format, etc., make sure that the mode is for multiple alignment.

Figure 12. Add sequence to CLUSTAL-X1.83

ii.   Multiple sequences alignment. Click 'do complete alignment', then there will be an interface for setting the memory way. There are two file formats, one is dnd which can be opened by treeview and another is aln in which the aligned sequences can be opened by CLUSTAL X and can be converted into MEGA file.
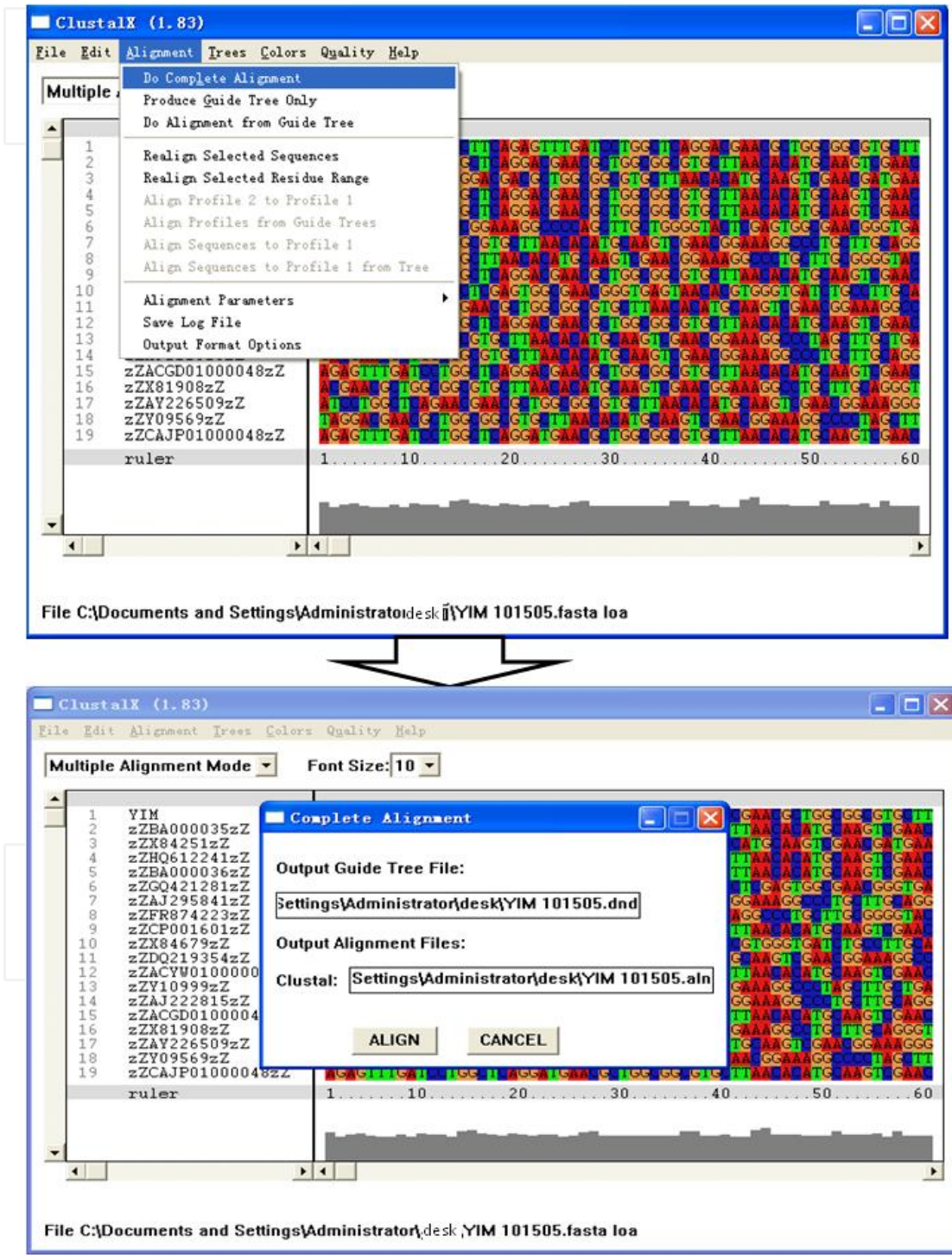

Figure 13. Align sequences by CLUSTAL_X1.83

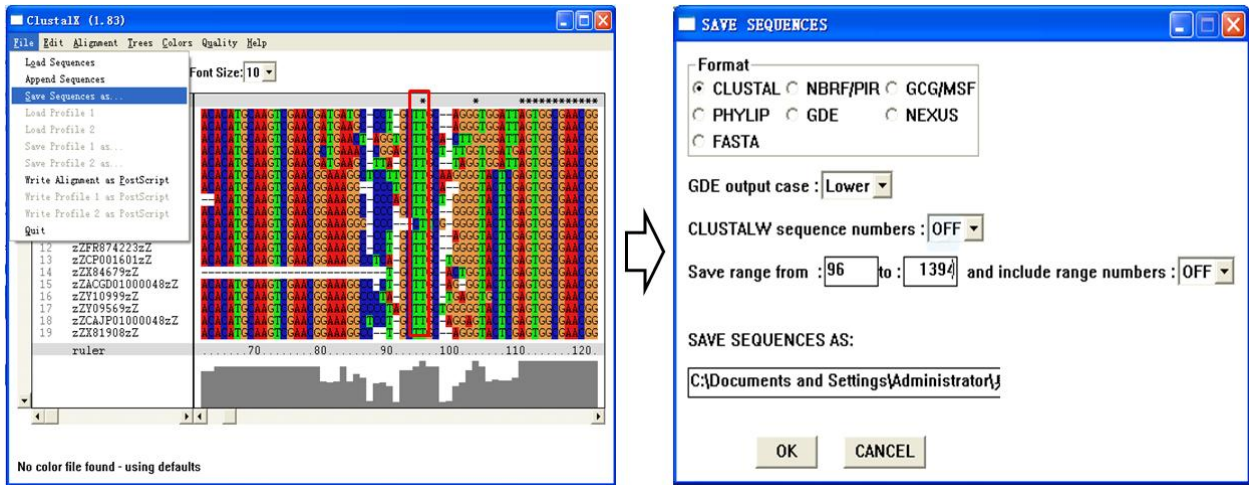**iii.**   Save sequence from the column with first '*' to the column with last '*' (see Figure 14. save range from 96–1,394) as file in clustal format.

Figure 14. Save aligned sequences from CLUSTAL_X1.83

iv.   Convert file in aln format into mega file format. Open MEGA6 and choose as shown in Figure 15..
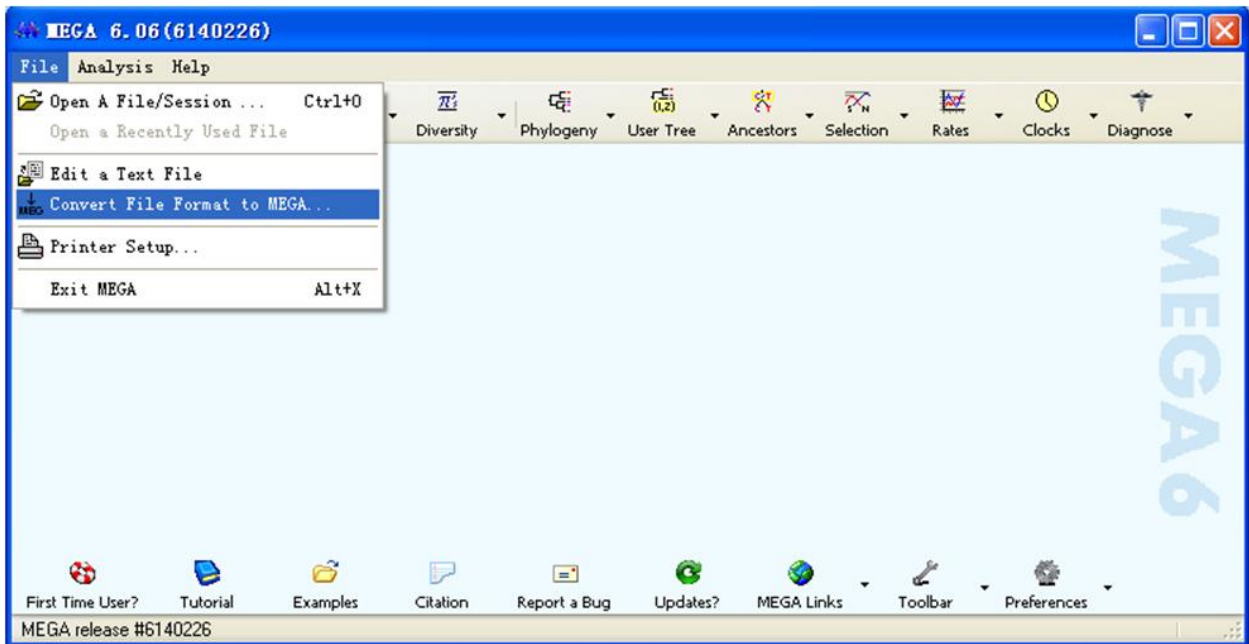

Figure 15. Convert file in aln format into mega file format

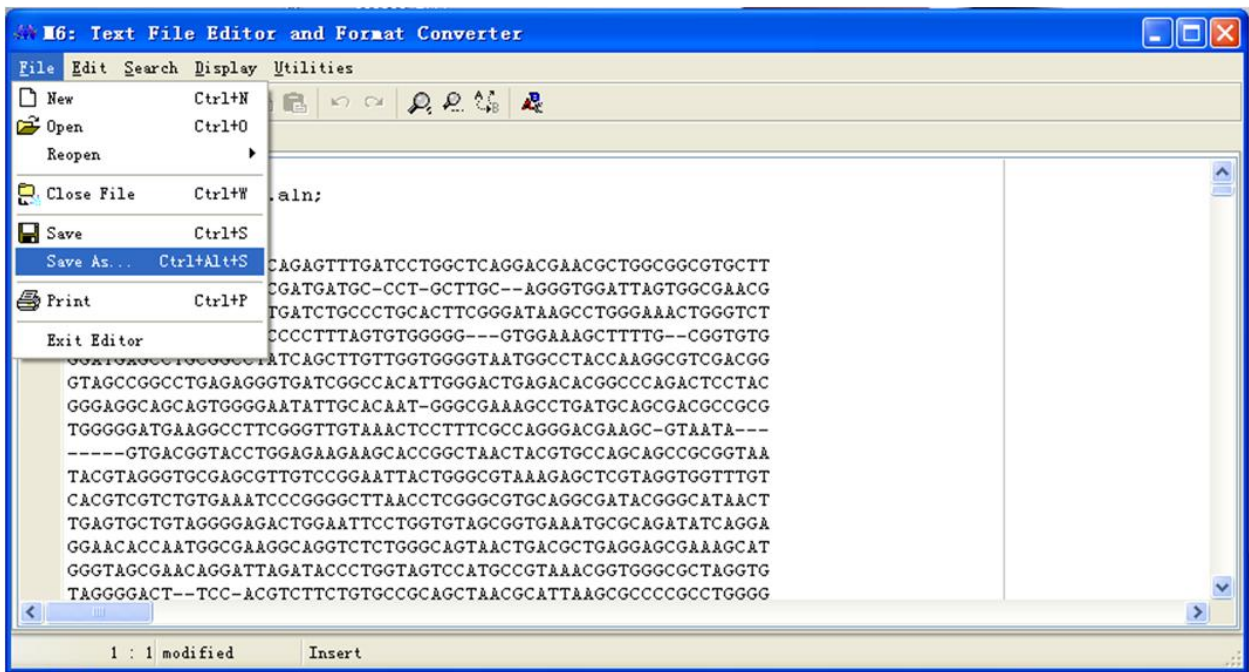**v.**   Save the converted file as a mega format (Figure 16).

Figure 16. Save the converted file

Sequences alignment by CLUSTAL W in the platform of MEGA6:

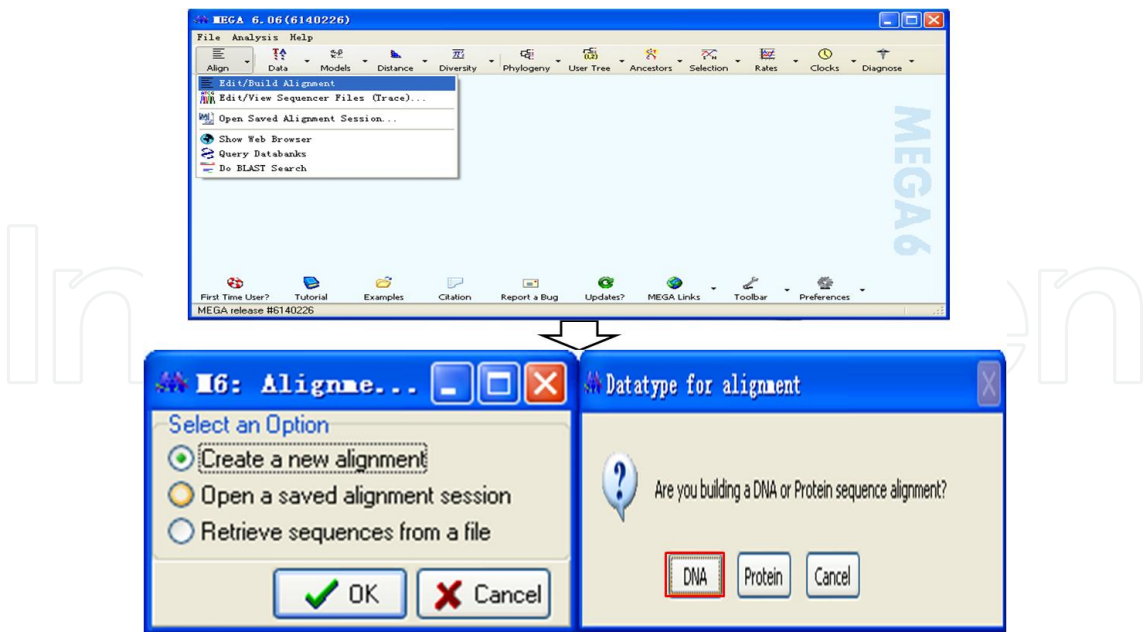**i.**   Open MEGA6 to build alignment as in Figure 17.

Figure 17. Build a DNA alignment by CLUSTAL W in MEGA6

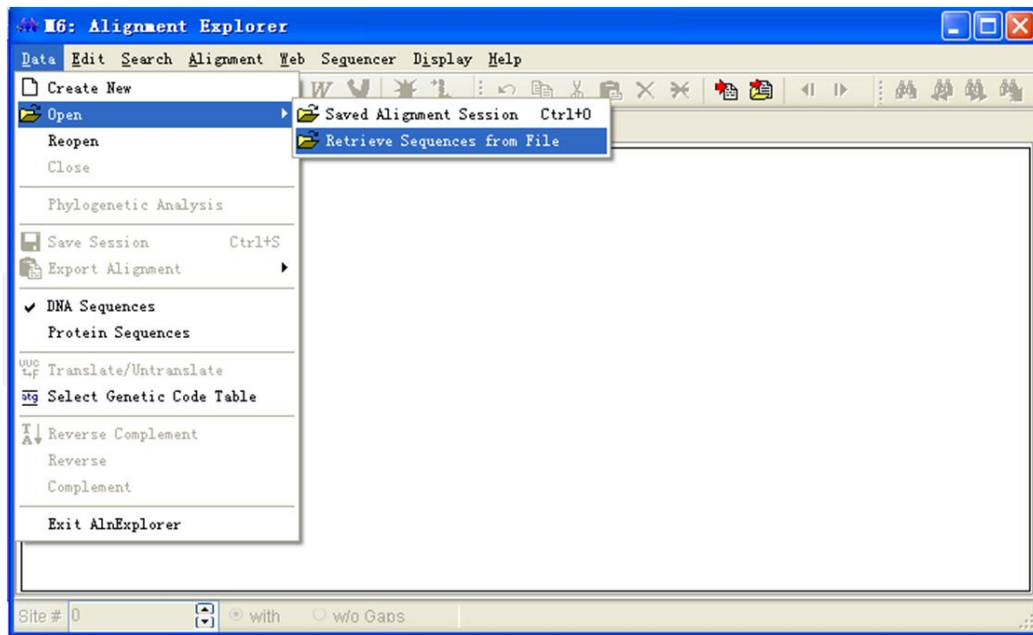**ii.**   Open a file from native computer as in Figure 18.

Figure 18. Add sequences to CLUSTAL W in MEGA6

iii. Multiple sequences alignment. Select all sequences and click 'align by ClustalW' to do alignment (Figure 19.). The parameters could be changed according to the illustration. Then there is an interface for setting the memory way. There are three file formats (mega\fasta\paup; Figure 20.); however, the mega format file is convenient for constructing a phylogenetic tree by MEGA6.
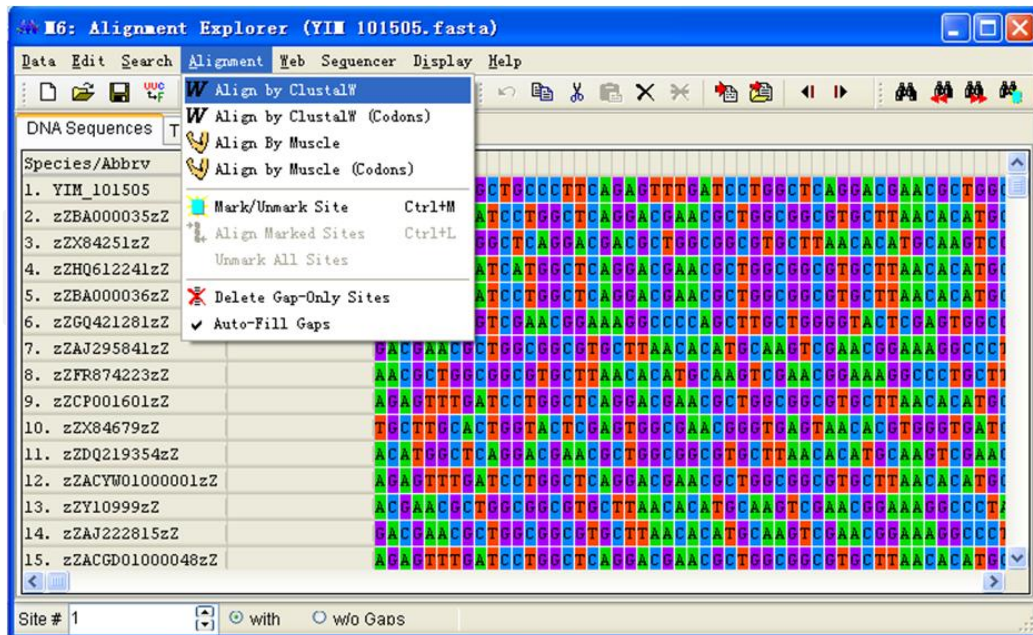

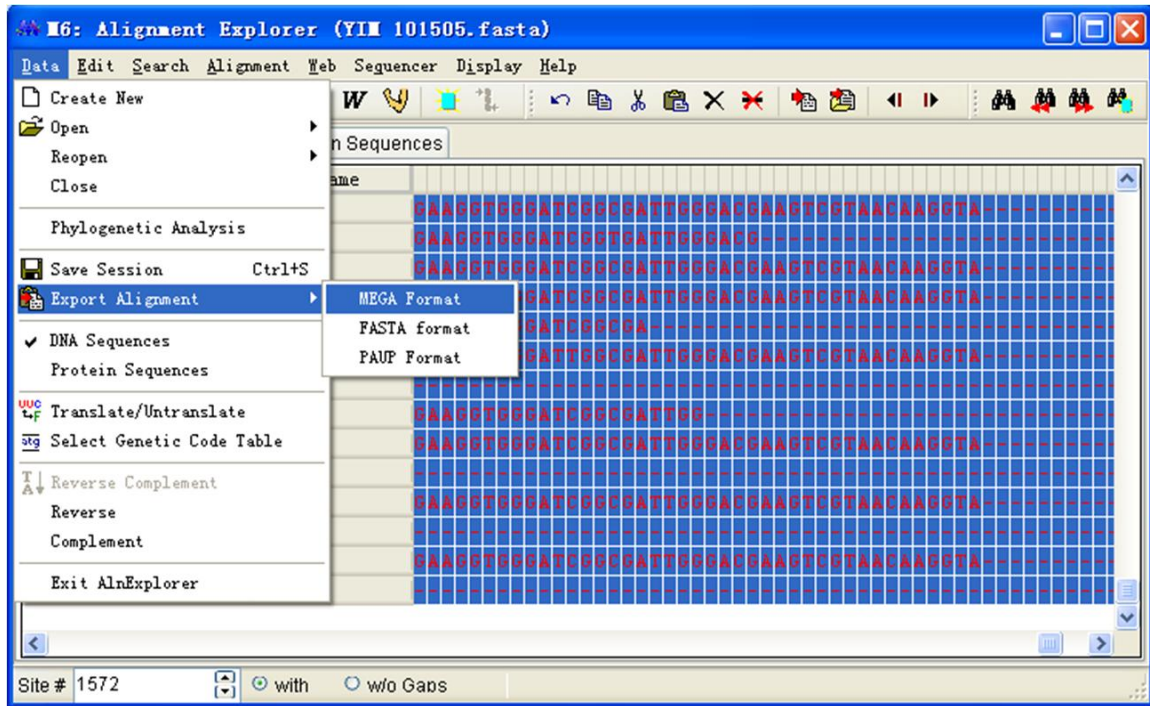**Figure 19.** Multiple sequences alignment

Figure 20. Save the result of alignment

4.3. Construction of phylogenetic tree by MEGA6

The Molecular Evolutionary Genetics Analysis (MEGA) software is developed for comparative analyses of DNA and protein sequences that are aimed at inferring the molecular evolutionary patterns of genes, genomes and species over time. It provides tree-making algorithms of maximum-likelihood, neighbour-joining, minimum evolution, UPGMA and maximum-parsimony. Bootstrap analysis is also included. In version 6.0, it added facilities for building molecular evolutionary trees scaled to time (timetrees), which are clearly needed by scientists as an increasing number of studies are reporting divergence times for species, strains and duplicated genes [24]. The following steps are used for construction of a neighbour-joining tree (some other tree can also be constructed following these steps when choosing different algorithms):

i.   Open MEGA6, click 'open a file' to activate a mega file. It is also available to open a mega file with MEGA6 directly. Choose nucleotide sequences in the input data and choose 'NO' to confirm for protein-coding nucleotide sequences data (Figure 21.).

ii. Click 'phylogeny' and then choose algorithms of neighbour-joining (Figure 22.).

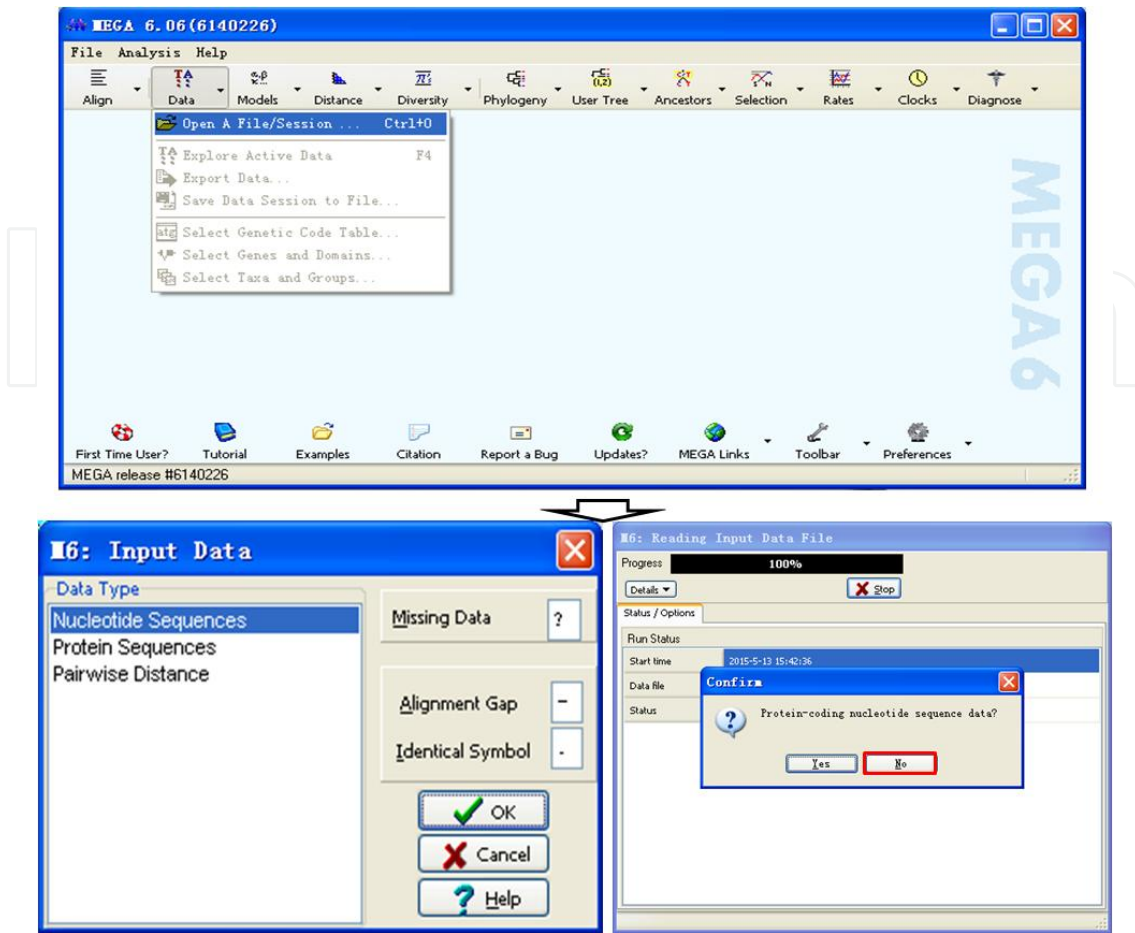iii. Analysis preferences are set as in Figure 23., then click 'compute' to get a phylogenetic tree (Figure 24).

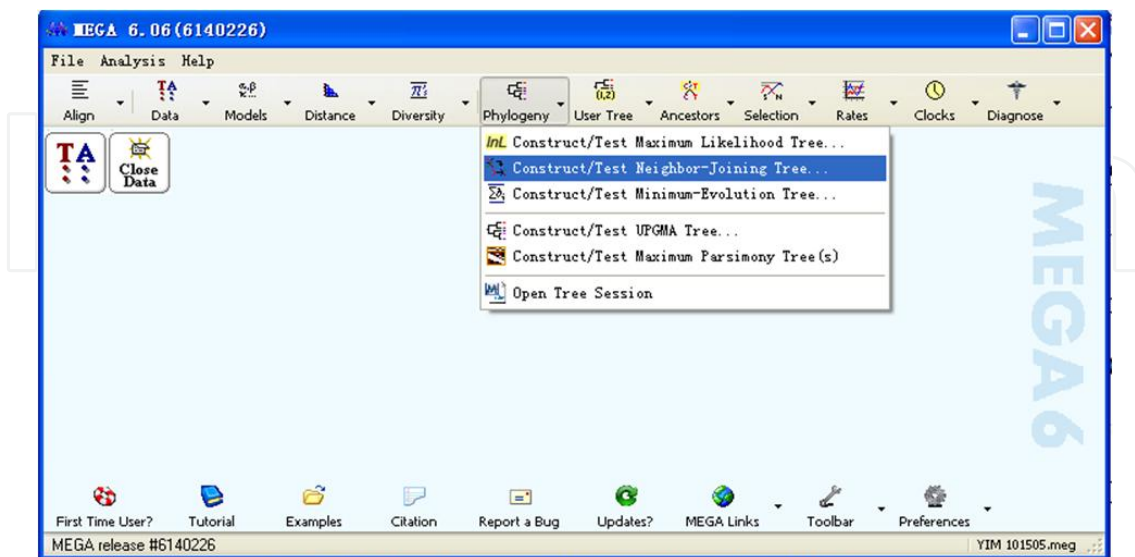**Figure 21.** Choices of constructing a phylogenetic tree based on DNA sequences
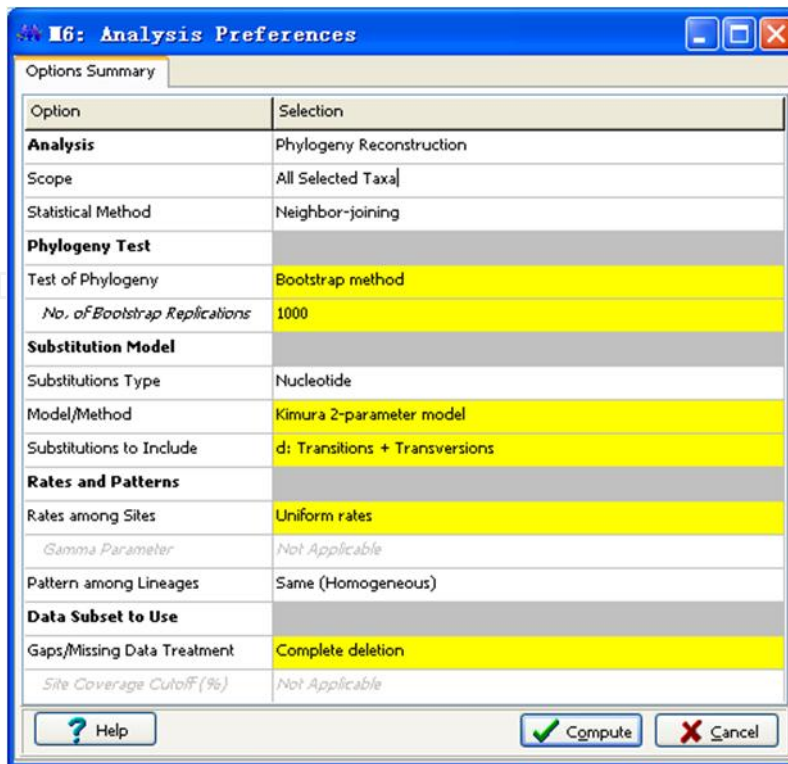
Figure 22. Algorithms in MEGA6

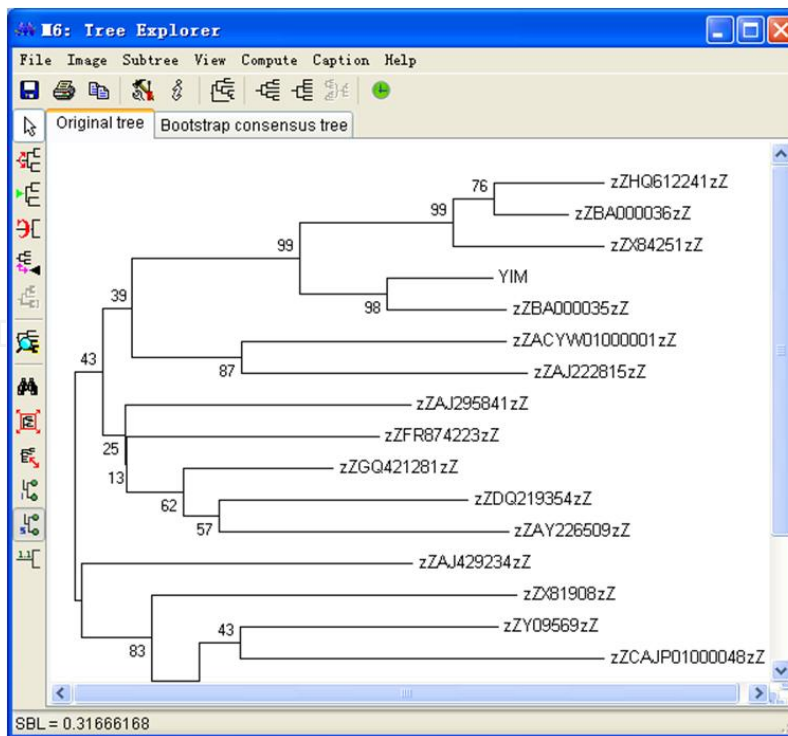**Figure 23.** Parameters for constructing a neighbour-joining tree

Figure 24. An example of a phylogenetic tree

iv.   Adjustment of the phylogenetic tree. The tree can be adjusted using buttons around the interface; the functions of some buttons are:

⇄, reverse the position of nodes of two clades.

⇄, reverse the position of two clades in one node.

◄, present subclade as a triangle.

, enlarge the subclade.

, definition of an outgroup for constructing a rooted tree.

, modify the shape of the tree, the width, length and bootstrap values, etc. could be set according to needs (Figure 25).
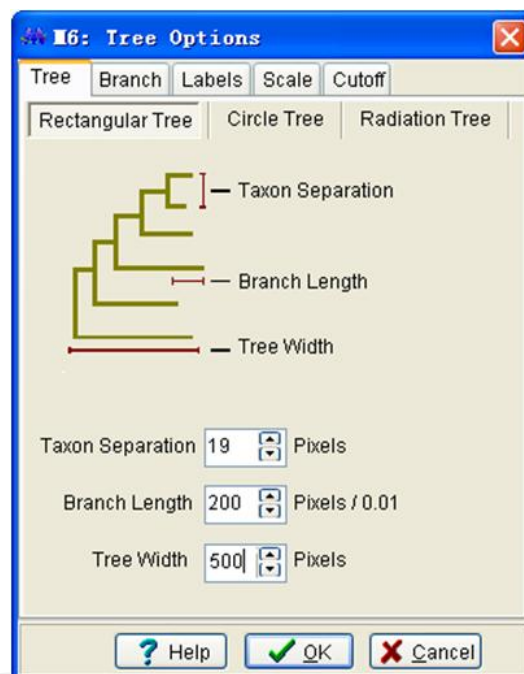
Figure 25. Interface for adjustment of tree

After the adjustment of the phylogenetic tree, save the tree in different formats or copy it into a file word format for edition.

The adjustment and editing of the phylogenetic tree in MEGA6 are limited, and there are different requirements for different journals, so the final edit of the tree is necessary. The following steps are for the edition of tree:

i.   Export current tree in Newick format (Figure 26.), and upload the file to http://www.ezbiocloud.net/eztaxon/replace_accession, then there will be an output of a file in the format of tree file which can be opened by MEGA6 and the accession on the branch is replaced (Figure 27.)

ii.   Copy the image to a file of.doc format, click the right click to edit the picture carefully.

Figure 26. Save tree as Newick format

Figure 27. Upload the tree as Newick format to replace_accession in Ezbiocloud

5. Determination of G+C content

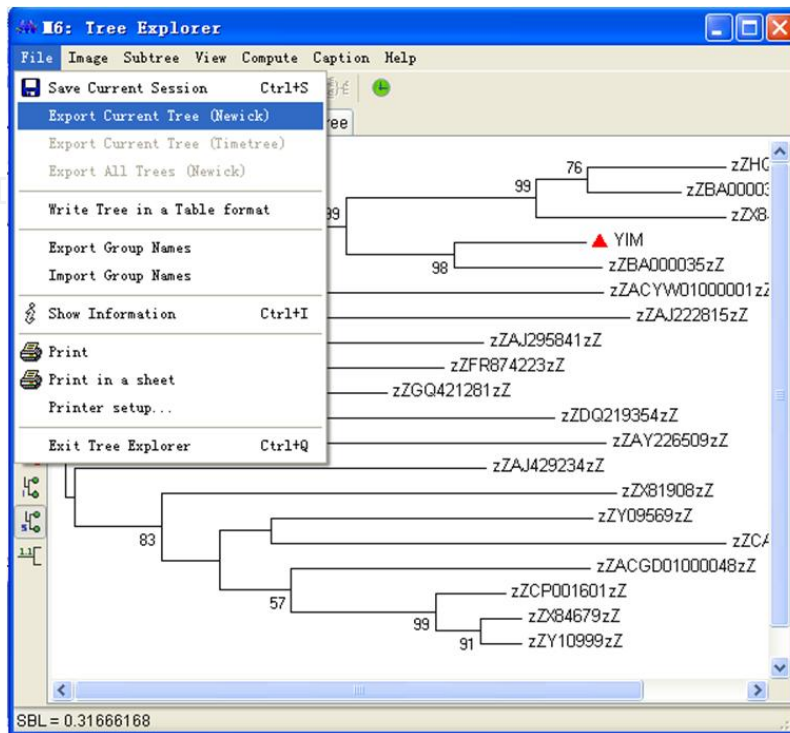The genomic DNA of each kind of organism has a specific G+C mol%, and G+C content varies in different organisms. Among the genotypic criteria for identification of bacteria, DNA G+C content has been widely used in bacterial taxonomy [13]. It is also an important prerequisite for determining the purity of DNA. The closer the two organisms, the more similar their G+C content is. However, the reverse of this reasoning is unreliable. Because G+C content of microbes is usually constant and not affected by age, growth condition and other external factors, so the determination of G+C mol% in the taxonomy and identification of microorganisms is of importance. The G+C mol% of most actinobacteria distributes between 50 and 80. The determination methods of G+C content are usually HPLC-based, although thermal stability of the native DNA and caesium chloride density-gradient centrifugation are alternative methods, these are now largely of historical interest [13]. HPLC-based is not affected by contamination with ribonucleic acid. Because this method yields a direct measurement, it may also be more accurate than indirect methods, such as the buoyant density and thermal denaturation methods. However, use of whole genome sequences to determine the G+C content of prokaryote will be more convenient in the future.

5.1. Determination of the G+C content of genomic DNA by High-Performance Liquid Chromatography (HPLC)

Escherichia coli should be performed as a control in using this method. Following steps are mainly referred to the method described by Mesbah [25].

i.   Extraction and purification of genomic DNA. The methods are detailed in section 6.2.

ii.   Determination of DNA concentration.

The absorption value at 280 and 260 nm is measured by using ultraviolet spectrophotometer, from which the purity and concentration of DNA can be determined. The value of OD_{260}/OD_{280} between 1.8 and 2.0 is qualified. The concentration of double-stranded DNA (μg/μl) between 0.1 and 1.0 is suggested.

iii.   Degradation of DNA.

    a.   Pipe 10 μl of a solution of DNA into a 200 μl microfuge tube and heated in PCR amplifier at 99℃ for 15 min. Take out the tube rapidly and cooled in an ice water.

    b.   Add 10 μl of P1 nuclease (63 U/ml in sodium acetate buffer; the buffer contains 2 mmol/l ZnSO4 and 40 mmol/l NaAc and the pH is 6.3), vortex gently. Then the sample is incubated for 2 h at 37℃ (the temperature can be improved, but the enzymic stability will decrease).

    c.   Add 10 μl of alkaline phosphatase (70 U/ml in Tris-HCl buffer; the pH of buffer is 8.3; In addition, the pH of the sample was between 7.5 and 8.5). Then the sample is incubated for 2 h (up to 6 h) at 37℃. Store the sample at –20°C for later use.

iv.   Determination of G+C content by HPLC.

The chromatography condition is listed in Table 2.

Chromatograph	Agilent 1100
Chromatographic column	(ZORBAX Eclipse XDB-C18) Analytical 4.6 × 150 mm 5-Micron
Mobile phase	0.05 mol/l $NH_4H_2PO_4$:C_2H_3N = 20:1
Detection wavelength	270 nm
Flow	1 ml/min
Column temperature	40℃
Injection volume	5–10 μl
Run time	10 min

Table 2. Chromatography condition of determination of G+C content

v.   Calculation of G+C content

G+C content can be calculated from the total component of DNA or from the ratio of certain bases. G+C content is defined as 100×M, where M is the mole fraction of deoxyguanosine (dGuo) plus deoxycytidine (dCyd). Thus, M = (G + C)/(G + C + A + T), where G, C, A and T are the mole fractions of the nucleosides dGuo, dCyd, deoxyadenosine (dAdo) and thymidine (dThd) (Figure 28.), respectively. When there is deviation for *E.coli*, the results should be revised. When unmodified bases are presented, G, C, A and T are the sums of the mole fractions of the modified and unmodified nucleosides.
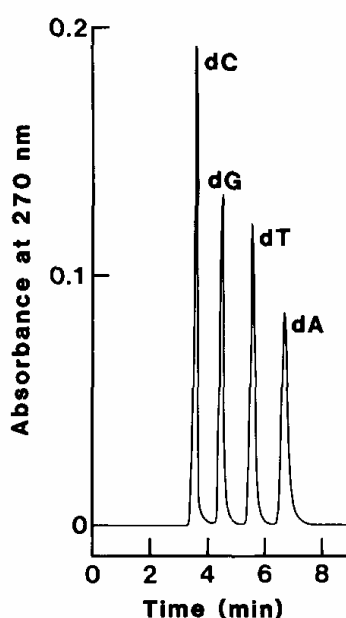

Figure 28. Four kinds of DNA nuclear nucleosides from standard HPLC chromatograms

6. DNA–DNA Hybridization (DDH)

DNA–DNA hybridization is one of the main procedures for identification of new species. Generally, DNA–DNA hybridization (DDH) is necessary when strains share more than 97% 16S rRNA gene sequence similarity. If a new research strain shows this high degree of similarity to more than one known species, DDH should be performed with all relevant type strains to ensure that there is sufficient dissimilarity to support the classification of the strain(s) as a new taxon. In 1987, the international system, International Committee on Systematic Bacteriology ICSB), provided that if the value of DDH is above 70% or the difference of melting temperature of hybrid molecular chain is less than 2℃, the two strains should be one species.

DDH can be performed using a number of techniques [13]. The first is liquid-phase DDH in which the hybridization reaction is in solution. The second is solid-phase DDH. The commonly used method of DDH is determined in micro-wells using covalent attachment of DNA. Total DNA for hybridization reactions is labelled with photoreactive biotin (photobiotin). The biotinylated DNA is hybridized with single-stranded unlabelled DNAs which had been immobilized on the surfaces of micro-dilution wells in this method. However, the DDH may be replaced by some genome relatedness indices, such as average nucleotide identity (ANI) [9], maximal unique matches index (MUMi) [26], genome BLAST distance phylogeny (GBDP) [27], and digital DDH (dDDH) which is computed using the recommended settings of the Genome-to-Genome Distance Calculator (GGDC) web server [28] version 2.0, etc. The protocol of DNA–DNA hybridization determined in micro-wells and DNA hybridization based on renaturation rates will be introduced in this part.

6.1. DNA–DNA hybridization determined in micro-wells

Ezaki *et al*. [5] compared the fluorometric hybridization method with a radioisotope method and firstly made it an alternative procedure to determine genetic relatedness among bacteria. Christensen *et al*. [6] made the modification by the addition of streptavidin conjugate alkaline phosphatase acting on the substrate 4-methylumbelliferyl phosphate in 2000. The protocol here mainly refers to the two articles. The salmon sperm DNA is performed as the control.

Extraction and purification of genomic DNA

i.   The methods are described in section 2. **Pretreatment of DNA (steps ii, iii)**

ii.   Diluted the DNA to 10 OD and 2 OD with 0.1×SSC.

iii.   Heat the DNA from step II at 100℃ for 15 min in tube, then transfer the tube to ice bath for 5 min immediately.

Binding of DNA to micro-wells (steps iv - viii)

iv.   Centrifuge at 1,000 rpm for 3 min, dilute the DNA (2 OD) to 0.2 OD with 1×PBS-$MgCl_2$.

v.   Add 100 µl DNA (0.2 OD from step IV) to each well.

vi.   Incubated micro-wells (with DNA) sealed with plastic bags at 30–50°C for a minimum of 4 h without shaking.

vii.   Wash the wells two times with 300 μl 1×PBS in one well each time.

viii.   Incubate the wells (with its original lid) at 40°C for about 30 min (until the well is dried).

DNA labelling with photo-activatable biotin (PAB), (under subdued light; step ix-xii)

ix.   Pipe 10 μl denatured DNA (10 OD from steps II and III) to mix with 10 μl PAB (prepare two tubes; salmon sperm DNA is included).

x.   Tubes are illuminated with their lids open, 10 cm below a 400 W Philips sun-lamp (SGR 140) for 90 min on crushed ice.

xi.   Add 200 μl TE buffer (pH 9) into the solution, vortex gently, and the solution is extracted twice (until the water phase is no longer to red) with 200 μl 2-butanol.

xii.   Shear the labelled DNA into fragments of 300–700 bp by ultrasonic wave (detect with agarose gel electrophoresis).

Pre-hybridization with unlabelled salmon sperm DNA to DNA attached to micro-wells (steps xiii– xv)

xiii.   Prepare pre-hydridization solution. 200 ml pre-hydridization solution including 40 ml 10×SSC, 20 ml 50×denhardt solution, 2 ml denatured salmon sperm DNA (10 mg/ml), 38 ml MilliQ water and 100 ml formamide.

xiv.   Add 200 μl DNA pre-hydridization solution (use it right after it was ready) in each well, sealed with plastic bags.

xv.   Incubate at hybridization temperature until the probe can add in (at least 60 min). The formula of TOR= 0.51 × (G+C mol%)+47, however, the TOR'=TOR-36+(0-5) when formamide is used.

Hybridization with PAB-labelled DNA to DNA attached to micro-wells (steps xvi–xxi)

xvi.   Mix 50 μl PAB-labelled DNA with 950 μl pre-hydridization solution to prepare hydridization solution, then incubate it for 10–15 min at hybridization temperature.

xvii.   Remove the pre-hybridization solution in the micro-wells completely.

xviii.   Add 100 μl DNA hydridization solution (use it right after it was ready) in each well, sealed with plastic bags and cover the micro-wells plate with silver paper, then incubate for at least 8 h at hybridization temperature.

xix.   Remove the hydridization solution in the micro-wells completely.

xx. Add 300 µl 1×SSC into each micro-well, incubate for 15 min at hybridization temperature, then remove the 1×SSC solution in the micro-wells completely for washing. Repeat three times.

xxi. Add 300 µl 1×SSC into each micro-well, incubate for 5 min at room temperature and then remove the 1×SSC solution from micro-wells completely for washing. Repeat three times.

Detection of DNA hybridization (steps xxii–xxvi)

xxii. Dilute streptavidin-conjugated alkaline phosphatase (VECTOR laborators) with alkaline phosphatase reaction buffer in a ratio of 1 : 3000.

xxiii. Add 100 µl mixture from step XXII into each micro-well, sealed with plastic bags and cover with silver paper, incubate for 1 h at 37℃.

xxiv. Wash the micro-wells with 100 µl alkaline phosphatase reaction buffer for each well, incubated for 5 min at room temperature every time. Repeat three times.

xxv. Add 100 µl 4-methylumbelliferyl phosphate (4-MUP) solution (1 mM; diluted with 4-MUP buffer) into each micro-well and which then be sealed with plastic bags and covered with silver paper, incubate for 1 h at 37℃.

xxvi. Fluorescence intensities were measured using a fluostar optima microplate reader (BMG LABTECH) at a wavelength of 360 nm for excitation and 460 nm for mission.

Quantification (step xxvii)

xxvii. The percentage DNA similarity was calculated as $100 \times [(I_{test} - I_{blank})] / [(I_{ref} - I_{blank})]$, where I_{test} is the intensity of hybridization between the strain to be tested and the reference strain, I_{ref}, is the intensity of hybridization of the reference strain with itself, and I_{blank} is the background hybridization (hybridization with salmon sperm DNA). Each experiment is performed with at least three replicates. The differences of mean DNA similarities between experiments are evaluated statistically by the d-test. The final similarity is the mean value of two independent experiments in which one is the DNA of tested strain as probe and another is the DNA of reference strain as probe.

6.2. DNA–DNA hybridization based on renaturation rate

This method needs a large amount of DNA and mainly according to the method described by De Ley *et al.* [29].

i. Extraction and purification of genomic DNA. The protocols are listed in section 2

ii. Dilute the DNA to 0.1 OD with 0.1×SSC.

iii.   Shear the genomic DNA into fragments of 200–1,000 bp (optimal 600 bp) by ultrasonic wave.

iv.   Denature NDA according to the set procedure. The procedure is set as Table.3

Procedure Number	Temperature/℃	Retention time/min	Run time/min
1-1	25	2	2
1-2	60	2	4
1-3	70	2	6
1-4	72	2	8
1-5	74	2	10
1-6	76	2	12
1-7	78	2	14
1-8	80	2	16
1-9	82	2	18
1-10	84	2	20
1-11	86	2	22
1-12	88	2	24
1-13	90	2	26
1-14	92	2	28
1-15	94	2	30
1-16	99	2	32
2-1	99	7	39
2-2	Tor*	2	41
2-3	Tor*	40	81
2-4	25	2	83

*Tor (Renaturation temperature)= 0.51 G+C mol% + 47.0

Table 3. Procedure of temperature and retention time

v.   Preincubate 20×SSC in boiling water.

vi.   Add 20×SSC to dilute the DNA following Procedure Number 1-16 to adjust the ion concentration, the final SSC concentration is 2×SSC.

vii. Determine the renaturation rate, obtain the data of resilience curve.

viii. Copy the data to Excel file, calculate the formula of velocity of renaturation based on the data above, and calculate renaturation velocity V.

ix. The DNA similarity is calculated as $100 \times [4V_m - (V_A + V_B)] / [(2 \times (V_A \times V_B)^{1/2})]$, V_A and V_B represent the renaturation velocity of sample A and B, V_m represents the renaturation velocity of mixture of sample A and B.

Acknowledgements

Author details

Xiu Chen[1,2], Yi Jiang[2], Qinyuan Li[2], Li Han[1] and Chenglin Jiang[1,2]

*Address all correspondence to: jiangyi@ynu.edu.cn

1 Institute of Microbial Pharmaceuticals, College of Life and Health Sciences, Northeastern University, Shenyang, P. R. China

2 Yunnan Institute of Microbiology, School of Life Science, Yunnan University, Kunming, P. R. China

References

[1] Lehmann KB and Neumann RO: Atlas und Grundriss Der Bakteriologie und Lehrbuch Der Speziellen Bakteriologischen Diagnostik. 1nd ed. Lehmann JF, Munchen, German), 1896.

[2] Orla-Jensen: Die Hauptlinien des naturalischen Bakteriensystems nebst einer Ubersicht der Garungsphenomene. *Zentralbl Bakteriol Parasitenkd Abt II*. 1909; 22: 305–346.

[3] Buchanan RE: Taxonomy. Annu Rev Microbiol. 1955; 9: 1–20. DOI: 0.1146/annurev.mi.09.100155. 000245

[4] Minnikin DE, Alshamaony L and Goodfellow M: Differentiation of Mycobacterium, Nocardia, and related taxa by thin-layer chromatographic analysis of whole-organ-

ism methanolysates. J Gen Microbiol 1975; 88: 200–204. Doi: 10.1099/00221287-88-1-200

[5] Ezaki T, Hashimoto Y and Yabuuchi E: Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in microdilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. Int J Syst Bacteriol 1989; 39: 224–229. doi: 0020-7713/89/030224-06$02.00/0

[6] Christensen H, Angen Ø, Mutters R, Olsen JE and Bisgaard M: DNA-DNA hybridization determined in micro-wells using covalent attachment of DNA. Int J Syst Evol Microbiol 2000; 50: 1095–1102. DOI: 10.1099/00207713-50-3-1095

[7] Gándara B, Merino AL, Rogel MA and Martínez-Romero E: Limited genetic diversity of Brucella spp. J Clin Microbiol 2001; 39: 235–240. DOI: 10.1128/JCM. 39.1.235-240.2001

[8] Coenye T and Vandamme P: Use of the genomic signature in bacterial classification and identification. Syst Appl Microbiol 2004; 27: 175–185. DOI: 10.1078/072320204322881790

[9] Konstantinidis KT and Tiedje J M. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microbiol 2007; 10: 504–509. DOI:10.1016/j.mib.2007.08.006

[10] Chun J and Rainey FA: Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int J Syst Evol Microbiol 2014; 64:316–324. DOI: 10.1099/ijs. 0.054171-0

[11] Stackebrandt E and Ebers J: Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006; 33:152–155.

[12] Stackebrandt E and Goebel BM: Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol 1994; 44: 846–849. DOI:10.1099/00207713-44-4-846

[13] Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W and Kämpfer P: Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 2010; 60: 249–266. Doi: 10.1099/ijs.0.016949-0

[14] Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M and Spratt BG: Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA 1998; 95: 3140–3145.

[15] Sullivan CB, Diggle MA and Clarke SC. Multilocus sequence typing: data analysis in clinical microbiology and public health. Mol Biotechnol 2005; 29:245–254. DOI: 10.1385/MB:29:3:245

[16] Tan ZY, Xu XD, Wang ET, Gao JL, Martinez-Romero E and Chen WX: Phylogenetic and genetic relationships of Mesorhizobium tianshanense and related rhizobia. Int J Syst Bacteriol 1997; 47: 874–879. DOI:10.1099/00207713-47-3-874

[17] Bouthinon D and Soldano H: A new method to predict the consensus secondary structure of a set of unaligned RNA sequences. Bioinformatics 1999; 15: 785–798. DOI: 10.1093/bioinformatics/16.10.785

[18] Akutsu T: Dynamic programming algorithms for RNA secondary structure predic‐ tion with pseudoknots. Discrete Appl Math 2000; 104:45–62. DOI:10.1016/ S0166-218X(00)00186-4

[19] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, Raoult D and Fournier PE: A polyphasic strategy incorporating genomic data for the taxo‐ nomic description of novel bacterial species. Int J Syst Evol Microbiol 2014; 64: 384– 391. DOI: 10.1099/ijs.0.057091-0

[20] Amaral GR, Dias GM, Wellington-Oguri M, Chimetto L, Campeão ME, Thompson FL and Thompson CC. Genotype to phenotype: identification of diagnostic vibrio phenotypes using whole genome sequences. Int J Syst Evol Microbiol 2014; 64: 357– 365. DOI: 10.1099/ijs.0.057927-0

[21] Chun J and Rainey FA: Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int J Syst Evol Microbiol 2014; 64: 316–324. DOI: 10.1099/ijs. 0.054171-0

[22] Meier-Kolthoff JP, Klenk HP and Göker M: Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol 2014; 64: 352– 356. DOI: 10.1099/ijs.0.056994-0

[23] Kerstin HE: Molecular Phylogenetic Analyses and Real Life Data. 2006. Universität zu Köln, Botanisches Institut, Lehrstuhl I, Gyrhofstr. 15, 50931 Köln, Germanye-mail: kerstin.hoef-emden@uni-koeln.de

[24] Tamura K, Stecher G, Peterson D, Filipski Aand Kumar S: MEGA6: Molecular Evolu‐ tionary Genetics Analysis version 6.0. Mol Biol Evol 2013; 30: 2725–2729. DOI: 10.1093/molbev/mst197

[25] Mesbah M, Premachandran U and Whitman WB: Precise measurement of the G+C content of deoxyribonucleic acid by high-performance liquid chromatography. Int J Syst Bacteriol 1989; 39: 159–167. DOI: 10.1099/00207713-39-2-159

[26] Deloger M, El Karoui M and Petit MA : A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. J Bacteriol 2009; 191: 91–99. DOI: 10.1128/JB.01202-08

[27] Henz SR, Huson DH, Auch AF, Nieselt-Struwe K and Schuster SC: Whole-genome prokaryotic phylogeny. Bioinformatics 2005; 21: 2329–2335. DOI: 10.1093/bioinfor‐ matics/bth324

[28] Auch AF, Klenk HP and Göker M: Standard operating procedure for calculating ge-
nome-to-genome distances based on high-scoring segment pairs. Stand Genomic Sci
2010; 2: 142–148. DOI: 10.4056/sigs.541628

[29] De Ley J, Cattoir H and Reynaerts A: The quantitative measurement of DNA hybridi-
zation from renaturation rates. Eur J Biochem 1970; 12:133–142. DOI: 10.1111/j.
1432-1033.1970.tb00830.x