Chapter 3

Materials and Methods
**Culturing of *Mycobacterium tuberculosis***

Middlebrook broth and solid media are popular culturing media for mycobacteria. Middlebrook minimal media is a mixture of salts which needs to be supplemented with oleic acid, bovine albumin, dextrose, catalase and sodium chloride (OADC Growth Supplement) for supporting bacterial growth. Extraneously added albumin can be a hindrance in proteomic experiment and hence 3 different albumin free media were tried – Sauton’s media, Praskauer-Beck media, and modified Reid’s synthetic medium. Optimal growth was obtained in Middlebrook 7H9 broth and Praskauer-Beck media. Middlebrook media and OADC supplements were obtained from HiMedia Laboratories. Praskauer-Beck media was prepared from individual components as follows -

**Monopotassium phosphate** 5.0g  
**Asparagine** 5.0g  
**Magnesium sulfate** 0.6g  
**Magnesium citrate** 2.5g  
**Glycerol** 20 ml  
**Distilled water** 1.0 L  

Above ingredients were dissolved in the same order making sure that each ingredient was dissolved before adding next one. pH was adjusted to 7.8 with NaOH. Mixture was autoclaved at 121°C for 15 minutes, after which pH was changed to 7.4. Medium was filtered to remove precipitate and autoclaved again at 121°C for 15 minutes.

*M. tuberculosis* H37Rv colonies from Lowenstein-Jensen media slants were used for inoculating 1 liter of liquid media (mentioned above). This reference strain is maintained at National JALMA Institute, Agra and National Tuberculosis Institute, Bangalore. Cultures
were grown at 37°C in stationary condition for 5 weeks. Growth was obtained in the form of floating as well as settled cell mass.

**Protein extraction from cell lysate**

For preparing cell lysate protein extract, culture from Middlebrook 7H9 media was used. At the end of five weeks, the cells were pelleted by spinning at 3000 rpm for 1 hour. Cell pellet was washed 3 times using chilled phosphate buffer saline. Cell lysis was carried out by bead beating (0.1 mm zirconia beads) in the presence of lysis solution. Zirconia beads were obtained from Biospec Products, USA. Bead beating was carried out in 1.5 ml microfuge tubes on Disuptor Genie from Scientific Industries Inc., USA. Protein from the cell lysate was estimated using Lowry Assay Kit (Bio-Rad Laboratories, USA). SDS (3 %) was used as lysis solution for SDS-PAGE fractionation and 9M urea was used as lysis solution for in-solution digestion. From ~3ml volume of wet cell pellet, 3-4 mg protein was obtained.

**Preparation of culture filtrate**

For preparation of culture filtrate proteins, *M. tuberculosis* culture from Proskauer-Beck media was used. This media was used as it does not contain added protein supplements (albumin and catalase). The cells were removed first by centrifugation at 3000 rpm for 1 hour and later by filtering through 0.22 µm membrane filter. Filtrate was concentrated using 3 kDa cut-off filters (Amicon). Protein estimation was carried out using Lowry assay kit (Bio-Rad Laboratories, USA).

**Sample preparation for mass spectrometry analysis**

With the aim of obtaining maximum proteome coverage, cell lysate was fractionated using 3 different strategies based on different properties of the peptide molecules. 1. SDS-PAGE followed by in-gel trypsin digestion; 2. Strong cation exchange (SCX) chromatography and
3. OFFGEL (IEF), the latter two being at the peptide level. Culture filtrate proteome is less complex than cell lysate proteome and considered to be an enriched fraction of secreted proteins. Hence, proteins were fractionated by SDS-PAGE alone (Figure 2).

**Figure 2.** Protein and peptide level fractionation strategy for deep proteome coverage

I) In-gel digestion of proteins

**Figure 3.** SDS-PAGE fractionation of *M. tuberculosis* cell lysate and culture filtrate
About 200 µg of cell lysate and culture filtrate protein extract was fractionated on 10% SDS-PAGE gel and stained using colloidal Coomassie blue stain (Figure 3). Excess stain was removed with the help of destainer. The cell lysate lane and culture filtrate lane was cut into 32 and 36 bands respectively. Bands were not uniformly cut but, each visibly discernible band was cut into a separate fraction. In-gel tryptic digestion and peptide extraction was carried out as described below -

1. The destained gel was rinsed with water and placed onto a clean glass plate. The lane of the protein sample of interest was separated out by cutting along the border. A clean scalpel blade was used to cut the visibly discernible bands.

2. Bands were excised into 1mm X 1mm pieces and transferred into micro-centrifuge tubes containing destaining solution (40mM ammonium bicarbonate- pH 8, 40% acetonitrile). Tubes were placed on a rocker/shaker and gently agitated to aid in destaining. The solution was discarded with a 200 µl pipette tip and the procedure was repeated until the gel pieces were completely destained.

3. Once the gel pieces were completely destained, tubes were spun and supernatant was discarded. Five hundred µl of 100% acetonitrile was added and incubated for 10-15 min until gel pieces were completely dehydrated and had become opaque.

4. Remaining acetonitrile was removed as much as possible with the help of gel loading tips.

5. Freshly prepared reduction reagent (5mM dithiothreitol in 40mM ammonium bicarbonate) was added to completely cover the pieces. Tubes were incubated at 60°C for 30 min.
6. Tubes were brought to the room temperature and reduction solution was removed from the tube with the help of 200 µl pipette tips. Freshly prepared alkylation reagent (20mM iodoacetamide (IAA) in 40mM ammonium bicarbonate) was added. Tubes were incubated in dark for 10 min (IAA solution should be prepared in amber colored or aluminum foil covered tubes).

7. Alkylating reagent was removed and gel pieces were dehydrated by adding 100% acetonitrile. Acetonitrile was removed after gel pieces were completely dehydrated and the tubes were kept on ice. Sequencing grade trypsin (Promega) dissolved in chilled 40 mM ammonium bicarbonate buffer was added to the tubes containing gel pieces. Amount of trypsin solution was decided based on number and size of gel pieces in the tubes.

8. Tubes were left on ice for 45 min until gel pieces have sufficiently imbibed the trypsin solution. More trypsin was added if necessary. Once the gel pieces were completely rehydrated, excess trypsin was removed and replaced with 40 mM ammonium bicarbonate to cover gel pieces. The tubes were incubated at 37°C overnight.

9. Peptide Extraction:- Tubes were cooled to room temperature and 100 µl of 5% formic acid was added. Tubes were spun and aqueous phase was collected in different set of tubes labeled according to fraction number. One hundred µl of extraction buffer (5% formic acid; 40% acetonitrile) was added to each tube and the tubes were kept on shaker for 30 min. Aqueous phase was collected in the tubes containing supernatant from the earlier step. This second step was repeated once or twice. Final extraction was carried out by adding 100% ACN. After gel pieces had completely dehydrated, supernatant was collected in the new tubes. Dehydrated gel pieces can be discarded.
10. Peptide extracts were dried down using vacuum dryer and stored for further analysis at -80°C. Dried peptides appear as translucent yellow to cream color film at the bottom of the tubes. If very less peptides were present no residue can be observed.

In-solution digestion of proteins

1mg of protein from cell lysate was used to make in-solution tryptic digest which was followed by peptide level fractionation. In-solution digestion was carried out as follows –

1. Protein concentration of cell lysate in 9M urea was determined by Lowry’s assay. Aliquot containing 1mg protein was taken in a 15 ml Falcon tube.

2. Reducing agent - dithiothreitol solution (50 mM) was added to the protein solution so that final concentration of dithiothreitol will be 5 mM. After gentle mixing the tube was kept at 60°C for 30 minutes.

3. After 30 minute incubation the tube was brought to room temperature and alkylating agent – iodoacetamide solution (100 mM) was added to the protein solution so that the final concentration is 10mM. The tube was kept in dark for 15 minutes.

4. An aliquot equivalent of 20 µg of protein was taken in a separate tube to be run on SDS-PAGE to check for protein integrity.

5. The reduced and alkylated protein solution was diluted with 40 mM ammonium bicarbonate so that final concentration of urea is reduced to 2 M from 9 M. This is done as trypsin action will be hindered in presence of 9 M urea.

6. Sequencing grade modified trypsin (Promega, USA) was added at the ratio of 1:20. That is, for 1,000 µg of protein 50µg of trypsin was added. Reaction was allowed to take place for up to 12-16 hours at 37°C.
7. At the end of reaction formic acid was added to the final concentration of 5% to stop the reaction. An aliquot equivalent to 20 µg of protein was taken out to run on SDS-PAGE to check for digestion efficiency (complete digestion is indicated by absence of bands after staining the gel).

8. The tryptic peptides obtained were desalted and stored at low temperature.

II) Cleaning/desalting of peptide preparation

Peptides were purified to remove any particulate matter or salts as it could hinder liquid chromatography process. For isoelectric focusing (OFFGEL) low ionic strength is desired. Desalting was done using Sep-Pak columns packed with reverse phase C_{18} material (Waters) as follows:-

1. Peptide solution (tryptic digest) was acidified to reach the pH range of 2-3 using 10% trifluoroacetic acid (TFA). The acidified solution was centrifuged at high speed (1000 rpm) to remove precipitate or particulate matter which can block the column.

2. C_{18} Sep-Pak column was fitted to 15 ml syringe. Column was pre-wetted using 100% acetonitrile. Solvent A (0.1% TFA) was passed through the column to remove acetonitrile.

3. Tryptic digest was loaded on top of the column and mobile phase was allowed to trickle down under gravity.

4. Column is washed with 12ml of solvent A, applied as 1 + 5 + 6 ml.

5. Peptides were eluted with Solvent B (0.1% TFA, 40% acetonitrile) applied as 3 + 3 ml.

6. Eluted peptides were vacuum dried and stored at -80°C.
III) Peptide fractionation using strong cation exchange (SCX) chromatography

Peptides equivalent to 350 µg of protein amount were fractionated by strong cation exchange (SCX) chromatography. Polysulfoethyl A column (PolyLC, 200x 2.1 5um, 200A) interfaced with Agilent 1100 HPLC system was used for SCX fractionation. Agilent HPLC system consisted of binary pump, auto-sampler, UV detector, fraction collector and external injector. The solvent system used was - solvent A - 10mM KH$_2$PO$_4$, 20% acetonitrile (pH 2.8) and solvent B - 10mM KH$_2$PO$_4$, 20% acetonitrile, 150mM KCL (pH 2.8). Peptides were dissolved in 1ml of solvent A and pH was adjusted to 2-3 using phosphoric acid. Peptides were loaded on SCX column with flow rate of 250µl/min. Fractionation was carried out by applying solvent B gradient as-

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>%B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>26</td>
</tr>
<tr>
<td>45</td>
<td>40</td>
</tr>
<tr>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>62</td>
<td>100</td>
</tr>
<tr>
<td>67</td>
<td>100</td>
</tr>
<tr>
<td>67.01</td>
<td>0</td>
</tr>
<tr>
<td>72</td>
<td>0</td>
</tr>
</tbody>
</table>

Fractions were collected every 1 minute. Elution profile was obtained by recording absorbance at 214 nm wavelength. A representative of typical tryptic peptide elution profile for *M. tuberculosis* is shown in Figure 4. Total 72 fractions were collected. Adjacent fractions were pooled based on the absorbance to get final 31 fractions to be analyzed by mass spectrometry. These fractions were also desalted using C$_{18}$ material packed in stage tips (procedure given on page 31).
IV) Peptide fractionation using OFFGEL fractionators (isoelectric focusing)

500 µg equivalent of peptides were fractionated using Agilent’s 3100 OFFGEL fractionators which fractionates peptide or proteins based on their isoelectric point. Twenty four fractions were collected by OFFGEL fractionation method over the pH range of 3 to 12. Fractionation was carried out as per the manufacturer’s instructions using reagents provided in the kit provided by manufacturer. Briefly, cleaned and dried peptides were dissolved in peptide OFFGEL stock solution containing glycerol and ampholytes to the final volume of 3.6 ml. Twenty four well frame was fitted on top of IPG strip. IPG strip was hydrated using 1.25 times diluted peptide OFFGEL stock solution. Peptide sample (150 µl) was loaded in each of 24 wells. The tray containing IPG strip and frame was place between 2 electrodes of the OFFGEL fractionator. Fractionation was carried out in constant voltage (500 Ohms) mode. At the end of fractionation, fractions were collected in separate tubes and cleaned as described below.
V)  **C$_{18}$ stage tip protocol for cleaning SCX and OFFGEL fractions.**

Individual SCX and OFFGEL fractions were cleaned/desalted using C$_{18}$ stage tips. Solvent A/washing solution was 0.1% TFA and solvent B/eluting agent was 0.1% TFA, 60% acetonitrile. Cleaning procedure followed was as follows -

1) C$_{18}$ stage tips were prepared by packing two ~2mm diameter discs punched out from 3M Empore C$_{18}$ discs in 10 µl micropipette tips. Pipetting needle (18X4") is used for punching out the discs and syringe attached at the back of the needle in used to release the punched discs into the 10µl tip. Gel loader tips or silica capillary is used to slide the material and fix it at the end of the tip.

2) Fractions are acidified by adding 10% TFA so that final concentration of TFA in the sample will be 1%.

3) C$_{18}$ material in stage tips in pre-wetted by 100% acetonitrile followed by washing with solvent A.

4) Acidified peptide sample is loaded on the C$_{18}$ material using Gel loader tip and pushed through the material using 10/15 ml syringe attached at the back of the stage tip.

5) Two washes of solvent A, each of 50 µl are given in the similar way.

6) Peptides are eluted by adding 30 to 50 µl solvent B and passing it through the C$_{18}$ material using syringe and slower speed. The stage tips can be incubated at room temperature for 10 minutes after pushing little bit of solvent B through the material to ensure efficient elution of the peptides.

7) Eluted peptides are vacuum dried and stored at -80°C until they were analyzed on mass spectrometer.
LC-MS/MS analysis

Total of 123 LC-MS/MS runs were carried out (Cell lysate in-gel fractions, 32; SCX fractions, 31; OFFGEL fractions, 24 and culture filtrate ingel fractions, 36). All of the mass spectrometry analyses were carried out on an LTQ-Orbitrap Velos ETD mass spectrometer (Thermo Scientific, Germany) interfaced with an Agilent 1200 series HPLC system. The peptides from each fraction were analyzed using reverse phase nano scale liquid chromatography (RPLC) coupled to tandem mass spectrometry. The RPLC system consisted of a 2 cm desalting column and a 10 cm long analytical column (75 µm x 10 cm, C$_{18}$ material 5 µm, 120 Å). Both these columns were packed in 75 µm silica tubing (Polymicro Tech, USA) with C$_{18}$ material of 5µm particle size and 120Å pore size (Michrome, USA) using pressure bomb. RPLC system was interfaced with mass spectrometer by an 8µm electrospray emitter tip (New Objective) maintained at 2.0 kV ion spray voltage.

The mass spectrometry analysis on the LTQ-Orbitrap Velos was carried out in a data dependent manner with survey scan resolution R= 60,000 at m/z 400, scan range of m/z 350 to 1,800. Up to 15 most abundant precursor ions were picked for MS/MS fragmentation by collision induced dissociation in each duty cycle (HCD mode was used for analysis of SCX and OFFGEL fractions, CID mode was used for in-gel fractions). Fragment ions were detected in Orbitrap with resolution R=15,000 at m/z 400. In the case of culture filtrate samples, fragment ion scans were acquired in the LTQ mass analyzer, where, resolution is >10 times lesser than that of Orbitrap analyzer. Further, ions picked for MS/MS were dynamically excluded for next 30 seconds. Normalized collision energy for MS/MS was set to 35%. Lock mass option was enabled which helps maintaining high mass accuracy by real time calibration using polysiloxane ions (445 m/z) from air.
Data analysis

I) Generation of peak list files

Raw data files (.raw files) were processed to generate peak list files. Mascot generic format (.mgf) files were generated using Proteome Discoverer software version 1.2 (Thermo Scientific, Germany). Following filtering parameters were used:

1) Allowed precursor mass range was 500 Da to 5000 Da.
2) Precursor charge state of 1 to 5 was allowed.
3) Minimum number of peaks in a spectrum was chosen to be 5.
4) Signal to noise ratio was set as 3.
5) For precursors with unrecognized charge state, default charge states of 2 and 3 were allowed.

II) Use of multiple search engines

To maximize the identification of peptides and proteins, three different search algorithms, Mascot (version 2.2), Sequest (SCM build 59) and MassWiz (version 1.6.4.3-A) were used to analyze the data. Mascot searches were submitted through Mascot Daemon to local Mascot server. Sequest searches were carried out using Proteome Discoverer software (Thermo Scientific). MassWiz searches were carried out on local server .mgf files were used as input for Mascot and MassWiz. .raw files were used as input files for Sequest.

III) Database searches for protein identification

Protein database used for MS/MS searches was downloaded from NCBI for Mycobacterium tuberculosis H37Rv strain (updated on April 12, 2009). The RefSeq database consisted of 3,988 protein sequences. Commonly encountered contaminant sequences were added to
protein database. Total number of sequences in protein database, including contaminants, was 4,015 Search parameters used were as follows-

a) Trypsin as a proteolytic enzyme allowing up to one missed cleavage.

b) Peptide mass error tolerance of 20 ppm

c) Fragment mass error tolerance of 0.1 Da (0.8 Da mass error was allowed for culture filtrate data as it was acquired in LTQ mass analyzer.)

d) Fixed post-translational modifications were carbamidomethylation of cysteine residues

e) Variable modifications allowed were oxidation of methionine, acetylation of peptide N-terminus and formylation of methionine.

A reverse sequence database was searched separately in addition to forward (target) database and false discovery rate (FDR) was calculated at every PSM score value as -%

\[
\% \text{ FDR} = \left( \frac{\text{Number of hits in reverse database at or above the score}}{\text{Total number of hits in target and reverse database at or above the score}} \right) \times 100
\]

A score threshold for 1% FDR was applied to search results from individual data files. Only first rank peptide sequences matching to each spectrum were considered. Since six different types of searches were performed with same data set (genome and protein database searches using Mascot, Sequest and MassWiz), peptide sequences assigned to a single spectrum in each search were compared. Spectra which were assigned different sequences in different searches were omitted from further analysis.

Protein identification list was generated by grouping proteins based on shared peptides. Proteins which have at least one unique peptide were selected from the group. From a group where no protein could be distinctly selected above others, that is, all proteins had equal evidence, it was represented by only one in the final list marked with an asterisk and
remaining equivalent protein IDs were reported in parentheses. Proteins with no unique peptide evidence (subset proteins) were reported in a separate list.

IV) Creation of alternative peptide sequence database for genome annotation

The genome sequence for H37Rv strain was downloaded from NCBI ftp site (NCBI Reference Sequence: NC_000962.2). Using in-house Python scripts, a six-frame translated database was created containing translated sequences from stop codon to stop codon. As *Mycobacterium* is known to use GTG and TTG as initiator methionine codons [6], wherever GTG and TTG codon was encountered it was translated as initiator methionine in addition to valine and leucine, respectively. The variant peptides thus obtained were appended to the genome translation databases as separate entries. Commonly encountered contaminants like bovine serum albumin (BSA), trypsin and keratins were added to the databases that were used for MS/MS ion search. Total number of sequences in genome translation database was 320,958. Mass spectrometric data obtained in high resolution mode at both MS and MS/MS levels were used for genome database searches and thus for proteogenomic analysis. This was done as we wanted to base our novel findings on high confidence peptide data only.

Search parameters used were as follows-

a) Trypsin as a proteolytic enzyme allowing up to one missed cleavage, semi-tryptic cleavage was allowed.

b) Peptide mass error tolerance of 20 ppm

c) Fragment mass error tolerance of 0.1 Da

d) Fixed post-translational modification was carbamidomethylation of cysteine residues.
e) Variable modifications allowed were oxidation of methionine, acetylation of peptide N-terminus and formylation of methionine.

Score thresholds for 1% FDR were calculated as described for protein database searches.

V) Workflow for genome annotation

Peptides obtained after applying 1% FDR cut off were selected for genome annotation analysis. Genome coordinates of all the peptides were found out using tblastn program. Peptides mapping to multiple places in the H37Rv genome were not considered for further proteogenomic analysis. Genome search specific peptides (GSSPs) were identified by excluding those peptides which mapped to known proteins. Next important step was to evaluate peptide spectral matches by manual validation. Criteria considered in manual validation of peptide spectral match are discussed in the following section.

Proteogenomic analysis pipeline adopted in the study is described in Figure 5. GSSPs were categorized as 1) mapping to intergenic region, 2) partially overlapping annotated genes, 3) completely mapping to annotated genes. Alternative gene models were searched in 10 kb region around the novel peptides identified. Two different gene prediction programs for prokaryotes were used – FgeneSB, GeneMark 2.5 [35, 36]. In addition, orfind tool (www.ncbi.nlm.nih.gov/gorf/) was also used where gene prediction programs did not produce alternative gene models. If the alternative gene model overlapped with existing gene model it was considered to be an instance of alternative gene structure or modification in the gene structure, otherwise it was considered to be a novel gene. Novel genes and gene structure modifications thus obtained using peptide evidence and gene prediction tools were checked
for their conservation across *Mycobacterium* spp. (in some cases *Corynebacterium* strains) using protein BLAST tool.

**Figure 5. Proteogenomic analysis pipeline.** Peptide qualifying 1% FDR threshold from both protein and genome search were considered for proteogenomic analysis. From six frame translated genome database search results, peptides which mapped to known proteins from protein database were eliminated to obtain a set of genome search specific peptides (GSSPs). GSSPs were classified as mapping to intergenic region partially overlapping with current gene annotation and completely mapping within current gene annotations which were further used to modify the present genome annotation. N-terminal peptides identified from protein database searches were used to either confirm or propose change in the annotated translational start sites.

*Kelkar et al., Mol Cell Proteomics. 2011. 10(12): M111.011627*
VI) Manual checking of peptide spectral matches

The novel peptides which passed 1% FDR score threshold were checked by manual inspection to check the quality of spectral assignment by the search engines. The genome search specific peptides (GSSPs) and novel N-terminal peptides which indicated alternative translational start sites were subjected to manual validation. The major criteria considered for manual evaluation included –

1) All intense peaks should have been assigned. Intense unassigned peaks were checked if they were arising from internal fragment ions.

2) Majority of the y series of ions should have been identified.

3) Low m/z range b ions, that is, b₁, b₂, b₃ ions and a₂ and a₄ ions were also observed in a typical spectrum.

4) The spectrum should have high signal to noise (S/N) ratio

5) If an immonium ion indicated the presence of an amino acid which was not present in the assigned peptide sequence, PSM was rejected.

6) Presence of y₁ ion confirming peptide ending either with K (m/z 147.11) or with R (m/z 175.12) was checked.

7) If the presence of any un-assigned fragment ion especially from higher m/z range indicated the presence of an amino acid which was not a part of assigned sequence, PSM was rejected.

8) Missed cleavage should be followed by either proline or acidic amino acid, that is, E and D.

9) If many assigned peaks were from the noise level then the PSM was rejected.

10) SOCH₄ neutral loss ions were observed for a peptide containing methionine.
VII) Use of Ensembl genome browser as visualization tool for proteogenomic analysis

Proteogenomic method used in this study was essentially a manual annotation strategy. A visualization platform like a genome browser forms an integral part of such analysis. Ensembl genome browser from the site (http://bacteria.ensembl.org) was used, where H37Rv genes as well as orthologs from different strains of Mycobacterium are mapped against *M. tuberculosis* H37Rv genome. As shown in Figure 6, good quality novel peptides were mapped on to the browser either using BLAST tools or as user uploaded tracks. The genomic sequence of interest within specific genome coordinates was obtained using the ‘export data’ utility from Ensembl browser.
Figure 6. Use of Ensembl genome browser for manual genome annotation. Top panel is a ‘region in detail’ panel where as middle and bottom panel are ‘region overview’ panel with zoomed out and zoomed in views respectively. Custom tracks were added (horizontal red bars) using coordinate files or BLAST tool. Translated sequence in six frames in zoomed in view helped in the detailed analysis of translational start and stop sites and frame changes.
Functional characterization of novel protein coding regions

Three different methods were used for this-

1) Protein BLAST- Functionally characterized homologous proteins were searched in non-redundant (nr) protein database using protein blast algorithm using novel protein sequence as query.

2) SMART domain analysis – Novel protein sequences were analyzed using SMART tools to identify known domains in SMART and pfam database [37].

3) Genomic context – In prokaryotic genome, genes belonging to the same operon are believed to have related function and are transcribed as single mRNA molecule. Potential operon structures/transcriptional units comprising novel genes were identified. Functional annotations of genes adjacent to the novel gene (genes at a distance of less than 200 bp and transcribed in the same orientation [38]) were borrowed to assign broad functional category to the novel gene.