Phylo- and comparative genomics of the *Pantoea* core genome

by

Marike du Plessis

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae (MSc)

to

The Faculty of Natural and Agricultural Sciences

University of Pretoria

Pretoria

Supervisor:	Prof. S. N. Venter
Co-supervisors:	Dr. M. P. A. Coetzee
	Prof. T. A. Coutinho
	Prof. E. T. Steenkamp

Declaration

I, the undersigned, hereby declare that the thesis submitted herewith for the degree Magister Scientiae to the University of Pretoria contains my own independent work. This work has not previously been submitted for any degree at any other University.

Marike du Plessis

June 2014

Acknowledgements

I would like to thank my supervisors, Profs. Fanus Venter, Emma Steenkamp, Teresa Coutinho and Dr. Martin Coetzee for all the guidance and patience they have shown during my degree. The advice and ideas that I have received from them were invaluable for my academic, as well as personal development. I would also like to acknowledge my funding bodies, the University of Pretoria (UP), the National Research Foundation (NRF), the South African Biodiversity Initiative (SABI), the Genome Research Institute (GRI) at the University of Pretoria and the Forestry and Agricultural Biotechnology Institute (FABI) together with the Tree Protection Co-operative Program (TPCP) and the DST/NRF Centre of Excellence in Tree Health Biotechnology (CTHB).

I would like to thank my family and friends for their continued support and love during this endeavour. Your support has been critical for the completion of this degree. A special thank you to Vernon Palmer for all the love, patience, support and understanding, I would not have succeeded without you.

Index

List of Tables

List of Figures

Preface

The delineation of bacterial species and genera has always been problematic as a clear definition of these concepts are lacking. In an attempt to classify bacteria into workable groups, operational criteria have been applied to delimitate boundaries for these taxonomic ranks. This approach has unfortunately led to artificial groupings that are often not comparable in terms of diversity in different groups of bacteria.

A classification system needs to reflect natural groupings to depict the evolution of bacteria and predict the phenotypic and genetic diversity for these groups. In order to understand the forces that play a role in the evolution of a bacterial genus a review of the current literature was presented in Chapter 1. The major focus was on vertical inheritance and how this process can be used to depict the evolutionary path of members belonging to the same genus. The largest amount of genetic material in any one cell is thought to have been transferred from parent to progeny, supporting the idea that the vertical signal is recoverable and can in fact be the dominant signal present in the genome when looking at conserved genes. The effect of horizontal gene transfer (HGT) on the evolutionary picture obtained by vertical descent was also discussed.

The core genome of a genus is defined as the genes conserved between all species of a genus and are thought to mostly include genes that are essential for the survival of members of that particular genus. In Chapter 2, the hypothesis was tested that the boundaries used to delineate genera could be based on an analysis of the shared core genome. For this purpose coherence within the core genome of the genus *Pantoea* was investigated. The core was characterised in terms of its functional diversity through Clusters of Orthologous Genes (COGs) and compared to the core genomes of other bacterial genera. It was seen that the core genome does give an indication of the coherence of a genus and that shared genome content can be used as a tool to delimitate genera.

Previous taxonomic studies have shown that species in the genus *Pantoea* are well defined but that the phylogenetic relationships between these species are not well elucidated. Generally accepted approaches for phylogenetic inference, like 16S rRNA gene trees and multi-locus sequence analysis (MLSA), does not give sufficient resolution to determine the deeper evolutionary relationships between these species. In Chapter 3, phylogenomic analyses were performed to determine if a robust phylogeny, reflecting the evolutionary history of the genus, can be obtained using the core genome of the genus. The core genome as well as subsets thereof (based on COGs), was used for phylogenetic inference, to obtain a robust phylogeny for the genus.

Chapter 1

Evolutionary forces that shape a bacterial genus

Evolutionary forces that shape a bacterial genus

Introduction

The sequencing of genomes has become an integral part of our endeavours to understand biology as it provides information on the genetic make-up and diversity of life on earth. Major breakthroughs in DNA sequencing technologies have provided researchers with data to understand and interpret biological functions and regulation, as well as the interactions and ecology of organisms (Hacker and Carniel, 2001). These advancements have led to the development of new fields such as synthetic and systems biology (Eisen and Wu, 2002; Brown and Sjölander, 2006).

The availability of genomes has made a major difference in the field of evolutionary biology and systematics. For the first time whole genome information can be used to characterize species and genera and to investigate their evolutionary relationships through comparative genomics and phylogenomics. As a result, several approaches to integrate genomics into the taxonomy of bacteria have been proposed (Rosselló-Móra, 2005; Richter and Rosselló-Móra, 2009; Thompson et al., 2013; Chun and Rainey, 2014). Through the construction of phylogenomic trees and networks, based on all the genes shared between organisms, both the vertical and the horizontal transfer of genes can also be determined (Bapteste and Burian, 2010). Genome sequence information thus facilitate the development of a more detailed understanding of the evolution of microorganisms and their associated characteristics, with the ultimate prospect of developing a natural classification system indicative of the evolution of the microorganisms (Klenk and Göker, 2010).

It is broadly accepted that bacterial species and genera, as they have traditionally been defined, often represent artificial groupings for circumscribing the diversity in clinical and industrial environments (Doolittle, 2009). The actual question is whether such discrete groupings exist in nature and are supported by the available genomic data. Indeed, the idea of a genetic continuum in terms of bacterial diversity has been strongly advocated based on the observed gene content variability within species and the abundance of horizontal gene transfer (HGT) amongst members of different species (Bapteste and Burian, 2010; Doolittle and Zhaxybayeva, 2010). These HGT events are characterized by the transfer of genes from one isolate to another through lateral means, thus not from parent to progeny, but through the acquisition of DNA from an external source (Andam and Gogarten, 2011).

In contrast to the genetic continuum idea, Caro-Quintero and Konstantinidis (2012) have used metagenomic data to show that discrete groups of diversity that are organised in genetically and ecological discernible populations exist within specific niches. Bennett et al. (2012) have also shown that reproducible species groups could be defined based on sets of shared core genes. Although some indication for the existence of discrete species exists based on genomic data, it is unclear whether similar groupings could be observed for the higher taxonomic ranks.

The aim of this literature review is to explore the forces that shape a bacterial genus and how this influences the systematics of bacteria. The use of different phylogenetic approaches for obtaining a species tree is discussed as well as the influence of horizontal acquisition of genetic material on bacterial evolution and the typical idea of vertical descent. Understanding how vertical descent and horizontal acquisition of genes influence bacterial evolution can provide a better idea of what comprises certain taxa. This may provide a better classification system that not only reflects a natural system, but also gives insights into the systematic of the organisms in terms of evolution and development.

Defining a bacterial genus

Compared to the delineation of bacterial species, defining bacterial genera has always been more problematic. The lack of a standard for describing taxonomic ranks higher than species level has led to the description of inconsistent groups with limited taxonomic predictive power (Konstantinidis and Tiedje, 2005b). Before the wide availability of molecular methods, bacterial genera were described solely based on morphological and physiological characteristics that were conserved between isolates (Wayne et al., 1987; Murray et al., 1990). The groupings obtained by only looking at phenotype, however, appear to be artificial (Konstantinidis and Tiedje, 2005b). Many characteristics considered as essential in these initial genus descriptions have been shown to not necessarily be present in all species belonging to the genus (Schubert, 1968). In other instances a single characteristic was used to unify species into a genus notwithstanding their highly divergent morphological and physiological characteristics. The genus *Hydrogenomonas* is a case in point. This genus was described based on the assumption that the utilization of H_2 by bacteria occurs infrequently enough to classify all of these species as members of the same genus (Davis et al., 1969). It was later shown that Gram-positive as well as Gram-negative bacteria were actually included in this genus and that some strains can lose this genus-defining characteristic when grown in culture (Davis et al., 1969).

Currently the main genetic criterion used for describing higher taxonomic ranks is that they should be supported by 16S ribosomal RNA (rRNA) gene phylogenies

(Konstantinidis and Tiedje, 2005b). This has led to taxonomic ranks that are not comparable between different groups. When whole-genome relatedness was compared to the current taxonomy, it was observed that many of the taxonomic ranks showed a high degree of overlap with adjacent ranks. This was especially the case for the ranks of genus, family and order (Konstantinidis and Tiedje, 2005b, Kunin et al., 2005), as many of the groups did not share the same level of genomic relatedness as determined by metrics such as average amino acid identity.

As molecular methods and sequencing data became more available, several attempts have been made to redefine some of the genera based on a combination of phenotypic, genotypic and phylogenetic data (Schubert, 1968; Gavini et al., 1989; Mergaert et al., 1993; Brady et al., 2010; Brady et al., 2013). This has led to the expectation of not only phenotypic cohesion, but also phylogenetic cohesion between species of the same genus. The current guiding principle often used is that when the available data suggests that a group of species is homogenous enough and sufficiently unique to identify them as different from the closest group, they can be defined as a separate genus (Gavini et al., 1989). As a result, numerous species have been moved from existing genera, to other or to newly defined genera (Gavini et al., 1989; Mergaert et al., 1993; Brady et al., 2010; Arora et al., 2011; Jung et al., 2012; Brady et al., 2013).

Defining bacterial taxa in terms of vertically inherited characteristics

In an attempt to develop a classification system that will reflect natural relationships and evolution (Gevers et al., 2005), a strong focus has been placed on the phylogenetic relationships between taxa. In order to recognise well-defined groups of isolates that are genetically distinct from their phylogenetic neighbours, it is necessary to trace the shared ancestry between these groups (Klenk and Göker, 2010). Here, vertical inheritance is emphasised, which, according to Andam and Gogarten (2011), "can be described as the evolutionary history of the majority of genes, as more genes will be inherited vertically in short periods of time than horizontally". The assumption is clearly that the dominant signal in the genome will be retained from the direct parental lineage. It also suggests that the signal from horizontal movement of genes will be overshadowed by the more dominant signal of vertical transfers and that a bifurcating tree will still depict the overall evolution of the taxa (Andam and Gogarten, 2011). The construction of 'species trees', therefore, appears to be the most likely approach to depict and understand the vertical history and relationships of bacterial taxa.

Over the past three decades, phylogenetic analyses based primarily on the 16S rRNA gene have been used to infer species trees for investigating the relationship between taxa (Konstantinidis and Tiedje, 2007). As ribosomes are absolutely essential for life, the small subunit genes are present in all organisms giving a universal marker (Hillis and Dixon, 1991), but also a marker generally considered to be vertically inherited. This gene has been widely used as different rates of evolution are observed across the gene giving different gene region options depending on the taxonomic level investigated (Hillis and Dixon, 1991). However, its level of divergence differs widely among taxa and in some groups it has remained much more conserved than the rest of their genomes (Konstantinidis and Tiedje, 2007). For most bacteria, 16S rRNA analysis can only discriminate accurately up to bacterial genus level (Konstantinidis and Tiedje, 2007) and is inadequate for differentiating between closely related bacterial species (Fox et al., 1992; Konstantinidis and Tiedje, 2005b; Staley, 2006; Richter and Rosselló-Móra, 2009;

Auch et al., 2010). An example of this is in the genus *Pantoea*, where 16S rRNA phylogenies cannot differentiate species reliably (Brady et al., 2008). For this genus 16S rRNA phylogenies also do not resolve its relationships with other closely related genera within the *Enterobacteriaceae* (Brady et al., 2008).

A lack of resolution is not the only problem experienced when using 16S rRNA data to determine taxonomic relationships. Phylogenies different from the evolutionary path can also be obtained from 16S rRNA sequence data. This has been observed in cases where the gene was subject to HGT (Boucher et al., 2004). Another issue when using 16S rRNA data is the sequence variation amongst multiple copies of the gene within the same genome (Boucher et al., 2004). This has been seen in some species of *Nocardia*, where at least three copies of the 16S rRNA gene, which differ in sequence, are present (Conville and Witebsky, 2007).

With the increased availability of gene sequences, multilocus sequence analysis (MLSA) has also been suggested for reconstructing the species tree (Gevers et al., 2005; Konstantinidis and Tiedje, 2007). MLSA is based on the analysis of a number of housekeeping genes (usually between 4 and 8) (Gevers et al., 2005; Konstantinidis and Tiedje, 2007) and has been widely applied for a range of bacterial groups (e.g., Kuhnert et al., 2009; Larsen et al., 2012; Karami et al., 2013; Octavia et al., 2013; Ramachandran et al., 2013). It is thought that the use of housekeeping genes for phylogenetic analyses can provide a picture of the evolutionary path followed by a group of organisms as these genes are usually essential in normal cellular functioning and the effects of HGT are thought to be less disruptive as multiple genes are analysed (Konstantinidis and Tiedje, 2007). The approach has proven useful for species identification and has provided some insight into the evolutionary relationships between taxa (Gevers et al., 2005; Konstantinidis and Tiedje, 2007). In *Pantoea*, for example, this approach is far superior to 16S rRNA analyses, and an MLSA tree based on only 4 housekeeping genes allows identification of species and provides far better resolution of the relationships between *Pantoea* and its closely related genera (Brady et al., 2008; Brady et al., 2009).

Constraints related to the use of MLSA, are similar to those observed for the 16S rRNA analyses. Housekeeping genes may remain more conserved than the rest of the genome, so that accurate separation between closely related groups cannot be achieved (Gevers et al., 2005). Horizontal transfer of these genes may also occur (Rivera et al., 1998) leading to topological differences between individual gene trees and the MLSA tree (Brady et al., 2008). Standardization of the genes used for MLSA studies is also problematic, as finding the same genes distributed throughout many different taxa is difficult and different gene sets are typically used for different phylogenetic groups (Gevers et al., 2005).

A recent extension of the MLSA is the use of all 53 ribosomal protein genes (Jolley et al., 2012). The products of these genes are proteins responsible for the scaffolding and rigidity of mature ribosomes (Hillis and Dixon, 1991). In a study where a number of isolates representative of the domain Bacteria were included, the use of this extended MLSA was shown to accurately identify isolates up to species level (Jolley et al., 2012). This approach has also been applied for the reclassification of *Neisseria* spp. (Bennett et al., 2012). A possible shortcoming may again be that multiple copies of some ribosomal protein genes have been observed in several bacterial isolates and HGT of certain ribosomal protein genes has been noted between various species (Boucher et al., 2004).

With the availability of whole genome sequences (WGS), large data sets for phylogenetic analysis (phylogenomics) have become available. Chan et al. (2012) used the genes shared amongst 10 *Acinetobacter* species to infer a species tree. Using a similar approach, Bennett et al. (2012) constructed a species tree for members of the genus *Neisseria* based on 246 shared genes. According to these researchers, the approach to use a large set of shared genes is basically also an extension of MLSA, with the advantage of bringing all informative characters into account (Chan et al., 2012). It therefore appears that the use of a set of shared conserved single-copy genes is currently the most reliable approach for inferring a species tree to determine the evolutionary relationships at a genus level (Lang et al., 2013).

The number of taxa and genes needed to infer a robust species tree

The main challenge, when constructing species trees, is to ensure that the tree based on the selected data set, truly reflects the overall evolutionary history of the taxa investigated. In an attempt to construct a reliable tree, various issues need to be considered, especially the number and type of taxa and genes.

It is generally thought that the robustness of the phylogenetic tree obtained increases with the addition of more taxa to the analysis (Kim, 1996; Graybeal, 1998; Mitchell et al., 2000). Hillis (1998) and others, however, suggested that increasing the number of taxa does not necessarily improve the accuracy of a tree when these taxa are highly divergent compared to the taxon or group in question (Kim, 1996; Mitchell et al., 2000). Taxon sampling will thus affect the accuracy of the tree when isolates are added to the tree which may not be informative. These highly divergent taxa may cause construction artefacts such as long-branch attraction, giving an inaccurate picture of the evolutionary path followed by the group of interest (Kim, 1996; Hillis, 1998; Mitchell et al., 2000). It is thus more reliable to focus sampling within a group to get more samples representative of the group of interest, rather than just adding more distantly related taxa that may cloud the picture (Hillis, 1998).

To increase the probability of obtaining the correct tree, it has been suggested that the number of genes to be used needs to be increased and not the number of alleles within a single species (Pamilo and Nei, 1988). This then suggests that if the number of genes is increased to all the genes shared between the isolates in question, the reliability of the tree will be increased. It is thought that the increased number of characters analysed, will help improve the signal to noise ratio, resulting in more reliable phylogenies (Klenk and Göker, 2010). There is, however, some evidence to suggest that uninformative genes do in fact lower support values for obtained phylogenies (Konstantinidis et al., 2006a; Salichos and Rokas, 2013) and that a few genes with strong phylogenetic signal are more useful than the full core genome data (Konstantinidis et al., 2006a; Salichos and Rokas, 2013). Although this may be the case, it is still believed that the core should be used during initial analyses to reflect the most probable vertical path of descent and could thereafter assist in identifying genes with strong signal for future use.

The selection of appropriate genes to be used for the construction of a species tree is critical and only orthologous genes should be brought into account. Paralogous copies of genes have a different evolutionary history compared to the rest of the genome and the phylogeny obtained may cause a skewed view of the relatedness that is not reflective of the true evolutionary history of the organisms. The detection of true orthologues can, however, be problematic (Coenye et al., 2005; Klenk and Göker, 2010). To avoid the issues related to gene duplication, phylogenetic analyses are often based only on single-copy genes.

Paralogous genes are common amongst bacteria as gene duplication has been described as a major evolutionary force in protein function (Kondrashov et al., 2002). It is generally thought that the presence of an additional copy of a gene, seen as redundant, allows the mutation and evolution of this extra gene, which leads to novel protein functions by releasing one of the copies from selective constraint (Ohno, 1970; Kondrashov et al., 2002). If the duplication was indeed redundant, fixation of the duplication will not occur. It has also been shown that in some cases relaxed selection could occur in both copies, which allows accelerated evolution of both copies (Kondrashov et al., 2002). Specialization in a bi-functional ancestral protein is an example (Kondrashov et al., 2002).

Core genome data for inferring the species tree

The term ‘species genome’ has been described as all the genes that can be found within at least one isolate of a species (Lan and Reeves, 2000). More recently this term has been replaced with the use of ‘pan genome’ (Tettelin et al., 2008; Mann et al., 2013), as this can apply not only to a species but also to higher taxonomic ranks. The pan genome of a group of organisms thus refers to all the genes found in at least one isolate within this group. This pan genome is further separated into a core genome and an accessory or auxiliary genome (Lan and Reeves, 2000; Tettelin et al., 2008; Mann et al., 2013). The core genome in its strictest sense, is defined as the genes that are present in all isolates within the group, and these genes are generally involved in essential cellular and biochemical functions necessary for survival (Lan and Reeves, 2000; Coenye et al., 2005; Hacker et al., 2012). The accessory genome consists of the remaining genes in the pan genome, thus genes that are unique to some isolates or shared by some but not all isolates in the group (Lan and Reeves, 2000; Coenye et al., 2005). These genes are thought to be associated with niche adaptation and specialization and are often located on genomic islands (Lan and Reeves, 2000; Hacker and Carniel, 2001; Coenye et al., 2005). The accessory genome is thus thought to be more prone to horizontal transfer, while the core genome is thought to remain conserved and passed along vertically (Daubin et al., 2002; Coenye et al., 2005).

A phylogeny inferred from core genome data will only trace the evolution of the shared genes among the taxa included. These core genes are usually restricted to the bacterial chromosomes and perform basic and essential cellular functions and generally have similar genomic features (Hacker and Carniel, 2001). Coenye et al., (2005) also suggested that sufficient phylogenetic signal may still remain in these genes to construct universal trees. However, this approach does not accommodate the evolution of the accessory genome (genes present in some but not all isolates) of a species/group of organisms (Doolittle and Papke, 2006). Bapteste et al., (2009) argue that this approach will result in a single species tree that may not accurately reflect the evolution of the species or genera involved (Bapteste and Burian, 2010), as it only focuses on a single aspect of the evolution of the isolates.

The core may represent the overall evolution of the taxa investigated as it has been suggested that the gene core in bacteria remains relatively cohesive (Daubin et al., 2002). Although bacteria lack sexual recombination, cohesion could be achieved through selective sweeps that may erase major genetic variation (Cohan, 2001; Cohan, 2005; Caro-Quintero and Konstantinidis, 2012). When adaptive mutations occur in individuals within a population, the individuals that have a competitive advantage may out-compete other individuals in the population, fixing the specific characteristic in that population (Cohan, 2001). This then allows the population in a specific niche to stay genetically similar. This has also been shown to be the case with adaptive evolution, where certain beneficial mutations become fixed within a population (Barrick et al., 2009), and although the population as a group may adapt, the population in that environment remains cohesive. Similarly, where specific populations of the same species are not isolated from each other indefinitely and get into contact with one another, the species as a whole may remain cohesive.

Some of the core genes used to infer the species tree may have different phylogenetic topologies or contradictory topologies compared to the gene trees. The incongruence of gene trees are most often due to genes being acquired through lateral means (Doyle, 1992; Galtier and Daubin, 2008) (see below). It is has been argued that phylogenetic trees are only valuable when recombination or HGT events are rare in the lineage/s of interest, otherwise the evolution of the lineage will not be accurately depicted (Bapteste and Burian, 2010). In contrast, others have argued that in general, recombination and HGT do not occur often enough between taxa to obscure the vertical signals used during the construction of phylogenetic trees (Cohan, 2001; Galtier and Daubin, 2008).

Horizontal acquisition of genes and the implications for bacterial systematics

Woese (2000) suggested that bacterial evolution would not be possible if both vertical as well as horizontal transmission of genes did not form part of the process.

This is consistent with the fact that HGT in prokaryotes has been observed more frequently than was initially thought (Boto, 2010), and has been identified as a major force in bacterial evolution (Gogarten et al., 2002; Jain et al., 2002; Cohen et al., 2010).

The acquisition of foreign DNA by bacteria can occur through conjugation, transformation and transfection (Hacker and Carniel, 2001; Jain et al., 2002; Boto, 2010). Conjugation is the process where two bacteria come together and DNA, typically plasmids, is transferred from the one bacterium to the other through a conjugation tube (Jain et al., 2002). The transfer of DNA from one cell to another via a virus is known as transfection (Jain et al., 2002). Transformation is the uptake of DNA from the environment by a bacterial cell (Jain et al., 2002), which may have originated from cells that have lysed and released their DNA into the environment. This uptake of free DNA is usually associated with bacterial species that are naturally competent, such as some *Bacillus subtilis* strains and other bacterial groups such as species in the genus *Neisseria* (Bennett et al., 2012).

HGT does not only relate to the transfer of single genes, but gene clusters or large genomic segments termed genomic islands, could also be transferred. These genomic islands are relatively easy to observe as they frequently harbour foreign elements like phage or plasmid elements, together with transfer genes and integration genes (Hacker and Carniel, 2001) and have different genomic features to the rest of the genome. The genes on genomic islands are grouped according to the functions of their products; e.g., some are involved in symbiosis (termed symbiotic islands), while others have a role in pathogenicity (termed pathogenicity islands) (Hacker and Carniel, 2001). The same genomic island may have a dual role, for example, in pathogenic bacteria present in the environment it may act as an ecological island, but once inside a host it may act as a pathogenicity island (Hacker and Carniel, 2001).

Horizontally transferred genes are generally considered as non-essential for basic survival and are associated with adaptation and confer specific beneficial attributes, which gives the bacterium a selective advantage (Hacker and Carniel, 2001). These genes could assist with the adaptation of the bacterium to changed environmental conditions within a specific niche (Gogarten et al., 2002; Coenye et al., 2005; Cohen et al., 2010). Another benefit may be that the acquisition of the new genes may assist the organism in invasion of an entirely new niche, where it did not previously occur (Gogarten et al., 2002). These genes may, for example, code for antibiotic resistance, pathogenicity factors or unconventional biochemical and metabolic features.

Although the transfer of genes is not exclusive to closely related organisms and can also occur between very distantly related organisms (Dagan et al., 2008; Boto, 2010), the observed frequency of transfer differ. When comparing closely related organisms to more distantly related organisms it has been shown that the HGT events occur more frequently between the closely related isolates (Andam and Gogarten, 2011). A possible explanation for this phenomenon may be that the cellular processes and genetic make-up of these organisms are similar, allowing these transfers to occur more readily (Andam and Gogarten, 2011). Such a process will allow related groups of organisms to remain cohesive as genetic material is constantly shared and exchanged within the group (Daubin et al., 2002; Andam and Gogarten, 2011).

Gene transfers have been found to occur more frequently between different species sharing the same niche or occupying overlapping niches (Gogarten et al., 2002; Andam and Gogarten, 2011). This may occur due to these organisms coming into contact with one another more frequently than with bacteria from different environments (Koonin et al., 2001). By populating a new niche, selective pressures may result in the uptake of related genes that will allow the strain to be better adapted to the new environment (Lan and Reeves, 2000; Gogarten et al., 2002). HGT is often responsible for the specialization of certain clones or the establishment of new clones with new abilities. The genetic variability of a species is thus often linked to the variety of niches that can potentially be inhabited by the different clones in a species and the potential horizontally obtained genes (Lan and Reeves, 2000).

Although most genes seem to be affected by HGT (Gogarten et al., 2002; Dagan and Martin, 2007; Dagan et al., 2008), there are genes that seem to be preferentially transferred (Cohen et al., 2010). There are generally two hypotheses with regard to the mechanism of HGT of transferred genes (Jain et al., 1999). The first hypothesis states that the transfer of genes has occurred continually between prokaryotes since the diversification of these organisms. The alternative hypothesis is that transfer of genes has occurred through the history of certain taxa as major events, where a large number of genes were transferred, with these events occurring rarely. Jain et al. (1999) have shown that the first hypothesis seems to be more likely as genes that are more prone to HGT tend to be incongruent with reference trees as well as between the individual gene trees, and does not form clusters of genes that remain congruent, as would be expected with major transfers occurring infrequently.

Rivera and colleagues (1998) differentiated genes according to function into informational genes and operational genes. It was shown that frequently transferred operational genes consists mainly of genes involved in metabolic capabilities and regulatory genes, while informational genes, like the genes involved in transcription, translation and other essential cellular processes, are transferred much less frequently (Rivera et al., 1998; Koonin et al., 2001). A possible explanation for this observed preference of certain genes in transfers was termed the complexity hypothesis by Jain and colleagues (1999). This hypothesis states that informational genes often occur in large groups of genes and that the gene products of these groups form very complex compounds that interact with one another, while the products of operational genes often function alone or needs less gene products to form functional complexes (Jain et al., 1999). Thus the transfer of informational genes may require multiple transfers of regions in the genome that are not contiguous, whereas the transfer of operational genes can occur independently without having an effect on the functionality (Jain et al., 1999). Cohen et al., (2010) have shown that this trend is due to the connectivity of these genes rather than the function. Therefore, the more connected protein families are through protein interactions, the less likely they are to be transferred horizontally.

As horizontal transfer of genes is thought to be primarily an adaptation mechanism (Gogarten et al., 2002), the majority of horizontally acquired genes are thought to be genes that form part of the accessory genome. If these genes do not confer an immediate beneficial phenotype, they cannot be beneficially involved in prokaryotic evolution as no selective advantage is given to the organism. Genes with limited beneficial contribution to the organism’s survival are easily lost (Lan and Reeves, 2000), as excess DNA is a burden and can be destructive to the cell (Doolittle and

Sapienza, 1980). Only when these transferred genes confer an advantageous attribute (Koonin et al., 2001), which ensures better fitness for the isolates possessing the gene, could this attribute potentially be fixed in the bacterial population (Koonin et al., 2001; Gogarten et al., 2002; Choi and Kim, 2007).

The above information suggests that it should be possible to obtain a tree depicting the probable vertical descent of the group using core genes that have been shown to be rarely transferred laterally. It has been suggested that if recombination and HGT amongst core genes occurred readily enough to eliminate the ability to track vertical descent, no sets of single gene trees with the same topology will be observed. This has, however, not been observed in phylogenomic studies (Klenk and Göker, 2010). It has also recently been shown that a core tree based only on single-copy genes, showing no signs of recombination or HGT, do not significantly differ from a tree constructed without filtering for these events (Chan et al., 2012). These observations thus support the notion that core genes could be used to construct a reliable species tree.

Considering all HGT events since the diversification of prokaryotes, the possibility exists that the majority of the genes present in organisms as we see them today may have been acquired laterally at some point (Gogarten et al., 2002; Dagan et al., 2008; Doolittle, 2009). However, these genes would have been passed along vertically afterwards and many will be part of the core genome of a specific taxon or group of taxa. If early gene duplications or HGT events occurred between deeper taxa, and these genes are present in all the progeny of a specific group, the transfers may have aided in diversification of the group and can still show a reliable traceable history of that specific group (Gao and Gupta, 2012). Such transfer events may, however, obscure the phylogenetic signals of higher taxonomic ranks (Klenk and Göker, 2010). This means that depending on the scale, the phylogenetic relationships among higher taxa could be blurred while it may still be possible to elucidate the relationships between closely related genera and species.

An opposing idea is that very little HGT is observed at higher taxonomic ranks. Very few transfer events could be detected by Choi and Kim (2007) between higher taxonomic ranks when horizontal transfer events were identified between 375 taxa. These taxonomic ranks included phylum, the range between phylum and order and the range between order and genus. At these taxonomic levels it was seen that horizontal transfer of genes are not widespread and that these events are not correlated to specific functions but rather occur continuously but to a low extent (Choi and Kim, 2007). This evidence supports the idea that HGT has occurred continuously and not as major events long ago (Jain et al., 1999) and that the effect of HGT would remain the same independent of the taxonomic level investigated.

As the full extent of HGT is not yet known, it has been suggested that it would only be useful to use a phylogenomic network or a sequence-shared network that can depict all interrelationships including horizontal transfers (Bapteste et al., 2009; Bapteste and Burian, 2010) to understand the relationship between bacterial strains. In phylogenomic networks, however, only extant organisms and DNA vectors are presented at the nodes, with no hypothetical ancestors being included. All nodes sharing a DNA sequence/s are linked in the network, showing the shared genes amongst each other (Bapteste and Burian, 2010), but not necessarily the vertical evolutionary path.

Conclusions

Despite the various difficulties associated with the inference of species trees, the concept of a species tree and the construction thereof, it still appears that a tree-like structure remains the most logical way to indicate phylogenetic relationships between organisms (Doolittle and Papke, 2006). Although the horizontal acquisition of genes may influence the classical idea of bacterial evolution through vertical descent, it further appears that sufficient vertical signal may remain in core genes to reconstruct a species tree. By obtaining a robust species tree, the origin of many group-specific characteristics, such as pathogenicity could be investigated. Although tree-like structures cannot bring the total evolutionary history of organisms into account, it can provide some idea of what bacterial species and genera are and how they have evolved giving a better idea of what bacterial systematics should reflect.

References

ANDAM, C. P. & GOGARTEN, J. P. (2011) Biased gene transfer in microbial evolution. *Nature Reviews Microbiology,* 9**,** 543-555.

ARORA, P. K., CHAUHAN, A., PANT, B., KORPOLE, S., MAYILRAJ, S. & JAIN, R. K. (2011) *Chryseomicrobium imtechense* gen. nov., sp. nov., a new member of the family Planococcaceae. *International Journal of Systematic and Evolutionary Microbiology,* 61**,** 1859-1864.

AUCH, A. F., VON JAN, M., KLENK, H.-P. & GÖKER, M. (2010) Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences,* 2**,** 117-134.

BAPTESTE, E. & BURIAN, R. M. (2010) On the need for integrative phylogenomics, and some steps toward its creation. *Biological Philosophy,* 25**,** 26.

BAPTESTE, E., O'MALLEY, M. A., BEIKO, R. G., ERESHEFSKY, M., GOGARTEN, J. P., FRANKLIN-HALL, L., LAPOINTE, F. J., DUPRE, J., DAGAN, T., BOUCHER, Y. & MARTIN, W. (2009) Prokaryotic evolution and the tree of life are two different things. *Biology direct,* 4**,** 34.

BARRICK, J. E., YU, D. S., YOON, S. H., JEONG, H., OH, T. K., SCHNEIDER, D., LENSKI, R. E. & KIM, J. F. (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature,* 461**,** 1243-1247.

BENNETT, J. S., JOLLEY, K. A., EARLE, S. G., CORTON, C., BENTLEY, S. D., PARKHILL, J. & MAIDEN, M. C. J. (2012) A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology,* 158**,** 1570-1580.

BOTO, L. (2010) Horizontal gene transfer in evolution: facts and challenges. *Proceedings of the Royal Society B: Biological Sciences,* 277**,** 819-827.

BOUCHER, Y., DOUADY, C. J., SHARMA, A. K., KAMEKURA, M. & DOOLITTLE, F. W. (2004) Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *Journal of Bacteriology,* 186**,** 3980-3990.

BRADY, C., CLEENWERCK, I., VENTER, S., COUTINHO, T. & DE VOS, P. (2013) Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA): Proposal to reclassify *E. nimipressuralis* and *E. amnigenus* into *Lelliottia* gen. nov. as *Lelliottia nimipressuralis* comb. nov. and *Lelliottia amnigena* comb. nov., respectively, *E. gergoviae* and *E. pyrinus* into *Pluralibacter* gen. nov. as *Pluralibacter gergoviae* comb. nov. and *Pluralibacter pyrinus* comb. nov., respectively, *E. cowanii*, *E. radicincitans*, *E. oryzae* and *E. arachidis* into *Kosakonia* gen. nov. as *Kosakonia cowanii* comb. nov., *Kosakonia radicincitans* comb. nov., *Kosakonia oryzae* comb. nov. and *Kosakonia arachidis* comb. nov., respectively, and *E. turicensis*, *E. helveticus* and *E. pulveris* into *Cronobacter* as *Cronobacter zurichensis* nom. nov., *Cronobacter helveticus* comb. nov. and *Cronobacter pulveris* comb. nov., respectively, and emended description of the genera *Enterobacter* and *Cronobacter*. *Systematic and Applied Microbiology,* 36**,** 309-319.

BRADY, C., CLEENWERCK, I., VENTER, S., VANCANNEYT, M., SWINGS, J. & COUTINHO, T. (2008) Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Systematic and Applied Microbiology,* 31**,** 447-460.

BRADY, C. L., VENTER, S. N., CLEENWERCK, I., ENGELBEEN, K., VANCANNEYT, M., SWINGS, J. & COUTINHO, T. A. (2009) *Pantoea vagans* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyi* sp. nov. and *Pantoea anthophila* sp. nov. *International Journal of Systematic and Evolutionary Microbiology,* 59**,** 2339-2345.

BRADY, C. L., VENTER, S. N., CLEENWERCK, I., VANDEMEULEBROECKE, K., DE VOS, P. & COUTINHO, T. A. (2010) Transfer of *Pantoea citrea*, *Pantoea punctata* and *Pantoea terrea* to the genus *Tatumella* emend. as *Tatumella citrea* comb. nov., *Tatumella punctata* comb. nov. and *Tatumella terrea* comb. nov. and description of

Tatumella morbirosei sp. nov. *International Journal of Systematic and Evolutionary Microbiology,* 60**,** 484-494.

BROWN, D. & SJÖLANDER, K. (2006) Functional classification using phylogenomic inference. *PLoS Computational Biology,* 2**,** e77.

CARO-QUINTERO, A. & KONSTANTINIDIS, K. T. (2012) Bacterial species may exist, metagenomics reveal. *Environmental Microbiology,* 14**,** 347-355.

CHAN, J. Z. M., HALACHEV, M. R., LOMAN, N. J., CONSTANTINIDOU, C. & PALLEN, M. J. (2012) Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*. *BMC Microbiology,* 12**,** 302.

CHOI, I.-G. & KIM, S.-H. (2007) Global extent of horizontal gene transfer. *Proceedings of the National Academy of Sciences,* 104**,** 4489-4494.

CHUN, J. & RAINEY, F. A. (2014) Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *International Journal of Systematic and Evolutionary Microbiology,* 64**,** 316-324.

COENYE, T., GEVERS, D., VAN DE PEER, Y., VANDAMME, P. & SWINGS, J. (2005) Towards a prokaryotic genomic taxonomy. *Federation of European Microbiological Societies Microbiology Reviews,* 29**,** 147-167.

COHAN, F. M. (2001) Bacterial species and speciation. *Systematic Biology,* 50**,** 513 - 524.

COHAN, F. M. (2005) Periodic selection and ecological diversity in Bacteria. IN NURMINSKY, D. (Ed.) *Selective Sweep.* Springer US.

COHEN, O., GOPHNA, U. & PUPKO, T. (2010) The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Molecular Biology and Evolution,* 28**,** 1481-1489.

CONVILLE, P. S. & WITEBSKY, F. G. (2007) Analysis of multiple differing copies of the 16S rRNA gene in five clinical isolates and three type strains of *Nocardia* species and implications for species assignment. *Journal of Clinical Microbiology,* 45**,** 1146-1151.

DAGAN, T., ARTZY-RANDRUP, Y. & MARTIN, W. (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proceedings of the National Academy of Sciences,* 105**,** 10039-10044.

DAGAN, T. & MARTIN, W. (2007) Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proceedings of the National Academy of Sciences,* 104**,** 870-875.

DAUBIN, V., GOUY, M. & PERRIÈRE, G. (2002) A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Research,* 12**,** 1080-1090.

DAVIS, D. H., DOUDOROFF, M., STANIER, R. Y. & MANDEL, M. (1969) Proposal to reject the genus *Hydrogenomonas*: taxonomic implications. *International Journal of Systematic Bacteriology,* 19**,** 375-390.

DOOLITTLE, W. F. & PAPKE, T. R. (2006) Genomics and the bacterial species problem. *Genome Biology,* 7**,** 116.

DOOLITTLE, W. F. & ZHAXYBAYEVA, O. (2010) Metagenomics and the units of biological organization. *BioScience,* 60**,** 102-112.

DOOLITTLE, W. F. (2009) The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 364**,** 2221-2228.

DOOLITTLE, W. F. & SAPIENZA, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature,* 284**,** 601-3.

DOYLE, J. J. (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany,* 17**,** 144-163.

EISEN, J. A. & WU, M. (2002) Phylogenetic analysis and gene functional predictions: Phylogenomics in action. *Theoretical Population Biology,* 61**,** 7.

FOX, G. E., WISOTZKEY, J. D. & JURTSHUK, P. J. (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *International Journal of Systematic Bacteriology,* 42**,** 166-170.

GALTIER, N. & DAUBIN, V. (2008) Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 363**,** 4023-4029.

GAO, B. & GUPTA, R. S. (2012) Microbial systematics in the post-genomics era. *Antonie Van Leeuwenhoek,* 101**,** 45-54.

GAVINI, F., MERGAERT, J., BEJI, A., MIELCAREK, C., IZARD, D., KERSTERS, K. & DE LEY, J. (1989) Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and Description of *Pantoea dispersa* sp. nov. *International Journal of Systematic Bacteriology,* 39**,** 337-345.

GEVERS, D., COHAN, F. M., LAWRENCE, J. G., SPRATT, B. G., COENYE, T., FEIL, E. J., STACKEBRANDT, E., VAN DE PEER, Y., VANDAMME, P., THOMPSON, F. L. & SWINGS, J. (2005) Re-evaluating prokaryotic species. *Nature Reviews,* 3**,** 733-739.

GOGARTEN, J. P., DOOLITTLE, W. F. & LAWRENCE, J. G. (2002) Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution,* 19**,** 2226-2238.

GRAYBEAL, A. (1998) Is it better to add taxa or characters to a difficult phylogenetic problem? *Systematic Biology,* 47**,** 9-17.

HACKER, J. R. & CARNIEL, E. (2001) Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO reports,* 2**,** 376-381.

HACKER, J. R. H., DOBRINDT, U. & KURTH, R. (2012) *Genome plasticity and infectious diseases*, ASM Press, Washington DC, USA.

HILLIS, D. M. (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology,* 47**,** 3-8.

HILLIS, D. M. & DIXON, M. T. (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology,* 66**,** 411-453.

JAIN, R., RIVERA, M. C. & LAKE, J. A. (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences,* 96**,** 3801-3806.

JAIN, R., RIVERA, M. C., MOORE, J. E. & LAKE, J. A. (2002) Horizontal gene transfer in microbial genome evolution. *Theoretical Population Biology,* 61**,** 489-495.

JOLLEY, K. A., BLISS, C. M., BENNETT, J. S., BRATCHER, H. B., BREHONY, C., COLLES, F. M., WIMALARATHNA, H., HARRISON, O. B., SHEPPARD, S. K., CODY, A. J. & MAIDEN, M. C. J. (2012) Ribosomal Multi-Locus Sequence Typing: universal characterisation of bacteria from domain to strain. *Department of Zoology.* Oxford, University of Oxford.

JUNG, M. Y., KIM, J.-S., PAEK, W. K., STYRAK, I., PARK, I.-S., SIN, Y., PAEK, J., PARK, K. A., KIM, H., KIM, H. L. & CHANG, Y.-H. (2012) Description of *Lysinibacillus sinduriensis* sp. nov., and transfer of *Bacillus massiliensis* and *Bacillus odysseyi* to the genus *Lysinibacillus* as *Lysinibacillus massiliensis* comb. nov. and *Lysinibacillus odysseyi* comb. nov. with emended description of the genus *Lysinibacillus*. *International Journal of Systematic and Evolutionary Microbiology,* 62**,** 2347-2355.

KARAMI, N., HELLDAL, L., WELINDER-OLSSON, C., ÅHRÉN, C. & MOORE, E. R. B. (2013) Sub-typing of extended-spectrum-β-lactamase-producing isolates from a nosocomial outbreak: application of a 10-loci generic *escherichia coli* multi-locus variable number tandem repeat analysis. *PLoS ONE,* 8**,** e83030.

KIM, J. (1996) General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Systematic Biology,* 45**,** 363-374.

KLENK, H. P. & GÖKER, M. (2010) En route to a genome-based classification of Archaea and Bacteria? *Systematic and Applied Microbiology,* 33**,** 175-182.

KONDRASHOV, F. A., ROGOZIN, I. B., WOLF, Y. I. & KOONIN, E. V. (2002) Selection in the evolution of gene duplications. *Genome Biology,* 3**,** 8.1-8.9.

KONSTANTINIDIS, K. T., RAMETTE, A. & TIEDJE, J. M. (2006a) The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 361**,** 1929-1940.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. (2005b) Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology,* 187**,** 6258-6264.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. (2007) Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Current Opinion in Microbiology,* 10**,** 504-509.

KOONIN, E. V., MAKAROVA, K. S. & ARAVIND, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification *Annual Reviews in Microbiology,* 55**,** 709-742.

KUHNERT, P., KORCZAK, B. E. M., STEPHAN, R., JOOSTEN, H. & IVERSEN, C. (2009) Phylogeny and prediction of genetic similarity of *Cronobacter* and related taxa by multilocus sequence analysis (MLSA). *International Journal of Food Microbiology,* 136**,** 152-158.

KUNIN, V., AHREN, D., GOLDOVSKY, L., JANSSEN, P. & OUZOUNIS, C. A. (2005) Measuring genome conservation across taxa: divided strains and united kingdoms. *Nucleic Acids Research,* 33**,** 616-621.

LAN, R. & REEVES, P. R. (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends in Microbiology,* 8**,** 396-401.

LANG, J. M., DARLING, A. E. & EISEN, J. A. (2013) Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices. *PLoS ONE,* 8**,** e62510.

LARSEN, M. V., COSENTINO, S., RASMUSSEN, S., FRIIS, C., HASMAN, H., MARVIG, R. L., JELSBAK, L., SICHERITZ-PONTÉN, T., USSERY, D. W., AARESTRUP, F. M. & LUND, O. (2012) Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. *Journal of Clinical Microbiology,* 50**,** 1355-1361.

MANN, R. A., SMITS, T. H. M., BÜHLMANN, A., BLOM, J., GOESMANN, A., FREY, J. E., PLUMMER, K. M., BEER, S. V., LUCK, J., DUFFY, B. & RODONI, B. (2013) Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. *PLoS ONE,* 8**,** e55644.

MERGAERT, J., VERDONCK, L. & KERSTERS, K. (1993) Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., respectively, and description of *Pantoea stewartii* subsp. *indologenes* subsp. nov. *International Journal of Systematic Bacteriology,* 43**,** 162-173.

MITCHELL, A., MITTER, C. & REGIER, J. C. (2000) More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). *Systematic Biology,* 49**,** 202-224.

MURRAY, R. G. E., BRENNER, D. J., COLWELL, R. R., DE VOS, P., GOODFELLOW, M., GRIMONT, P. A. D., PFENNIG, N., STACKEBRANDT, E. & ZAVARZIN, G. A. (1990) Report of the ad hoc committee on approaches to taxonomy within the Proteobacteria. *International Journal of Systematic Bacteriology,* 40**,** 213-215.

OCTAVIA, S., SALIM, A., KURNIAWAN, J., LAM, C., LEUNG, Q., AHSAN, S., REEVES, P. R., NAIR, G. B. & LAN, R. (2013) Population structure and evolution of non-O1/non-O139 *Vibrio cholerae* by multilocus sequence typing. *PLoS ONE,* 8**,** e65342.

OHNO, S. (1970) *Evolution by gene duplication*, London: George Alien & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag.

PAMILO, P. & NEI, M. (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution,* 5**,** 568-583.

RAMACHANDRAN, P., LACHER, D. W., PFEILER, E. A. & ELKINS, C. A. (2013) Development of a tiered multilocus sequence typing scheme for members of the *Lactobacillus acidophilus* complex. *Applied and Environmental Microbiology,* 79**,** 7220-7228.

RICHTER, M. & ROSSELLÓ-MÓRA, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences,* 106**,** 19126-19131.

RIVERA, M. C., JAIN, R., MOORE, J. E. & LAKE, J. A. (1998) Genomic evidence for two functionally distinct gene classes. *Proceedings of the National Academy of Sciences,* 95**,** 6239-6244.

ROSSELLÓ-MÓRA, R. (2005) Updating prokaryotic taxonomy. *Journal of Bacteriology,* 187**,** 6255-6257.

SALICHOS, L. & ROKAS, A. (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature,* 497**,** 327-333.

SCHUBERT, R. H. (1968) The taxonomy and nomenclature of the genus *Aeromonas* Kluyver and van Niel 1936 Part III. Suggestions on the definition of the genus *Aeromonas* Kluyver and van Niel 1936. *International Journal of Systematic Bacteriology,* 18**,** 1-7.

STALEY, J. T. (2006) The bacterial species dilemma and the genomic-phylogenetic species concept. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 361**,** 1899-1909.

TETTELIN, H., RILEY, D., CATTUTO, C. & MEDINI, D. (2008) Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology,* 11**,** 472-477.

THOMPSON, C., CHIMETTO, L., EDWARDS, R., SWINGS, J., STACKEBRANDT, E. & THOMPSON, F. (2013) Microbial genomic taxonomy. *BMC Genomics,* 14**,** 913.

WAYNE, L. G., BRENNER, D. J., COLWELL, R. R., GRIMONT, P. A. D., KANDLER, O., KRICHEVSKY, M. I., MOORE, L. H., MOORE, W. E. C., MURRAY, R. G. E., STACKEBRANDT, E., STARR, M. P. & TRÜPER, H. G. (1987) Report of the Ad Hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic Bacteriology,* 37**,** 436-464.

WOESE, C. R. (2000) Interpreting the universal phylogenetic tree. *Proceedings of the National Academy of Sciences,* 97**,** 8392-8396.

Chapter 2

Characterizing the core genome of a genus: the case of Pantoea

Characterizing the core genome of a genus: the case of *Pantoea*

Introduction

In bacterial systematics, species delineation is often problematic (Doolittle and Papke, 2006). Although this is probably a consequence of not having an appropriate species concept based on sound biological and evolutionary principles (Gevers et al., 2005; Doolittle and Papke, 2006; Thompson et al., 2009), the situation is exacerbated by lateral and environmental forces that influence the evolution of bacteria (Doolittle and Papke, 2006). For example, the effects of horizontal gene transfer (HGT), as well as niche adaptation and specialization may significantly blur species boundaries (Coenye et al., 2005). To circumvent these complications, bacterial species are defined based on operational criteria that are thought to reflect phenotypic, phylogenetic and genomic similarities (Cohan, 2001; Gevers et al., 2005; Richter and Rosselló-Móra, 2009; Thompson et al., 2009). This approach, however, has many limitations and does not give consistent results across all bacterial taxa (Gevers et al., 2005; Konstantinidis et al., 2006; Goris et al., 2007).

The delineation of bacterial genera has been an even greater challenge for taxonomists as the concept of what comprises a genus is not clearly defined or understood. In contrast to species, only limited guidelines for the delineation of bacterial genera currently exist (Konstantinidis and Tiedje, 2005a; Tindall et al., 2010). Apart from 16S rRNA gene sequence similarities (Konstantinidis and Tiedje, 2005a), a limited number or diverse range of characters are used as support for groupings (Winslow et al., 1917). These characters typically include colony morphology in terms of pigmentation and appearance, general growth conditions and nutritional requirements and physiologic and metabolic capabilities (Winslow et al., 1920; Whitman et al., 2012). However, all of these characters have limited capacity to predict evolutionary relationships and often result in artificial groupings (Konstantinidis and Tiedje, 2005a; Konstantinidis and Tiedje, 2005b). In many cases, some of these genus-defining characters are also absent from some members of the same genus (Schubert, 1968; Brady et al., 2010b).

The dramatic developments in sequencing technologies and the extensive use thereof in biology has benefited bacterial systematics (Gevers et al., 2005; Thompson et al., 2009; Klenk and Göker, 2010; Gao and Gupta, 2012) and it is now possible to study the relatedness of bacteria using whole genome sequence (WGS) data. At present WGS data is primarily used to investigate evolutionary relatedness (Ciccarelli et al., 2006; Comas et al., 2007; Edwards et al., 2007), while metrics such as average amino acid identity (AAI), average nucleotide identity (ANI), genomic signatures and codon usage bias are used to assess the level of genetic and phylogenetic cohesion among organisms (Konstantinidis and Tiedje, 2005b; Konstantinidis and Tiedje, 2007; Richter and Rosselló-Móra, 2009). Although a set of genome-based guidelines have been proposed for the delineation of bacterial species (Konstantinidis and Tiedje, 2005a; Konstantinidis and Tiedje, 2006; Richter and Rosselló-Móra, 2009; Thompson et al., 2011), the results of a number of studies have shown that the delineation of genera using WGS data remains complex. For example, studies utilizing AAI showed that the genetic relatedness among several species currently grouped in the same genus is lower than expected and that the species in question should rather be grouped as belonging to the same family or order (Konstantinidis and Tiedje, 2005b; Kunin et al., 2005). These studies thus highlight the fact that our understanding of a bacterial genus requires significant improvement.

The availability of WGS data has allowed for the comparison of gene content across individual isolates, species and genera by using pan genome analyses focusing on both the accessory (or ancillary) and core genomes (Lan and Reeves, 2000; Tettelin et al., 2008; Mann et al., 2013). The accessory genome consists of the genetic material unique to some isolates or genes not shared by all isolates of the taxon (Lan and Reeves, 2000; Coenye et al., 2005). The products of these genes may be involved in diverse processes ranging from general metabolism through to pathogenicity and virulence (Coenye et al., 2005), which may facilitate survival in certain environments. The core genome, on the other hand, consists of those genes present in nearly all the members of a taxon (Lan and Reeves, 2000; Coenye et al., 2005) and usually represent only a small subset of a group’s pan genome (Lukjancenko et al., 2012). Genes included on the core genome are considered to be essential for survival and often encode products involved in crucial cellular processes (Hacker et al., 2012). The core genome also provides information on the shared functions and characteristics of bacterial groups (Lan and Reeves, 2000).

The phylogenetic information captured on the core and accessory genomes are markedly different. Analysis of the accessory genome is thought to primarily provide information on the ecology or lifestyle of organisms (Lan and Reeves, 2000; Hacker and Carniel, 2001; Coenye et al., 2005) and how these organisms have adapted to their specific environments. Contrast to the accessory genome that is subject to significant levels of HGT, the core genomes of species and genera are typically used to reconstruct their evolutionary histories (Daubin et al., 2002; Coenye et al., 2005) as the core genes may collectively represent the dominant evolutionary signal that depicts vertical descent (Daubin et al., 2002).

The way in which the core genomes of bacteria evolve suggests that this genomic component may be used to delineate the boundaries of genera. Evaluation of this hypothesis is dependent on the availability of whole genome sequence (WGS) data for a well-defined and comprehensively sampled genus. Previous functional characterization and gene content analyses of core genomes have focussed mainly on economically important species (Loper et al., 2012; Lukjancenko et al., 2012; Mann et al., 2013). However, the genus *Pantoea* is well suited as a model for addressing this hypothesis as it is phylogenetically cohesive and includes a range of well-defined species (Brady et al., 2007; Brady et al., 2008) that are ecologically and phenotypically diverse. Numerous *Pantoea* species are associated with plants where they are pathogens (Brady et al., 2008; Brady et al., 2009; De Maayer et al., 2012a, 2012b; Kim et al., 2012) or plant-growth promoters (Kim et al., 2012; Montañez et al., 2012), some are human pathogens (Brady et al., 2010a) or can infect humans opportunistically (De Baere et al., 2004; Cataño et al., 2012), and yet others are associated with insects (Adams et al., 2011; Kobayashi et al., 2011). Whole genome sequence information for a number of genomes representative of the genus are available.

In this study, the hypothesis that the core genome of a genus can be used to delineate its boundaries was evaluated by making use of the available WGS data for *Pantoea*. The specific aim was therefore to determine whether the *Pantoea* core genome will reflect the expected intra-generic cohesion and thus allow delineation of the genus. For this purpose the putative core genome for the genus *Pantoea* was determined from the publicly available genome sequence data using EDGAR (Blom et. al., 2009), after which it was characterized in terms of its general cellular processes and functions based on Clusters of Orthologous protein Groups (COGs)

(Tatusov et al., 2000). The putative core genome of *Pantoea* was also compared to the core genomes of various other genera from the Proteobacteria, as well as the Firmicutes, to determine possible similarities in size, functional diversity and AAI. It is believed that this information and insights will be invaluable for defining genus boundaries for *Pantoea* as well as other bacterial groups in general.

Materials and Experimental methods

Determining the core genomes and COG associations of core genomes

The core genome of *Pantoea* was compared to those of *Rhizobium (*including *Agrobacterium), Burkholderia, Ralstonia, Erwinia, Escherichia, Shewanella, Listeria* and *Lactobacillus*. These genera ranged from closely related to distantly related genera and included some genera from the classes α-, β- and γ-Proteobacteria, as well as two genera from the class Bacilli (The phylogenetic relationship of these genera is indicated in Figure 1). For *Pantoea, Ralstonia* and *Erwinia*, core genome sizes were also determined when a representative of their respective sister genera had been included. These calculations were determined to evaluate the difference in core genome size of a genus relative to a genus plus its sister taxon. For the comparison with the *Pantoea* core genome, the *Tatumella morbirosei* LMG 23360^T genome was used. For the *Erwinia* core genome comparison, the *Pantoea ananatis* LMG 20103 genome (De Maayer et al., 2010) was used. For the *Ralstonia* core genome comparison, the genome sequence of *Cupriavidus taiwanensis* LMG 19424^T (Amadou et al., 2008) was used.

The core genome genes were downloaded for all genera considered in this study from the EDGAR (Efficient Database framework for comparative Genome Analyses using BLAST score Ratios) database (http://edgar.cebitec.uni-bielefeld.de/cgi-bin/edgar_login.cgi; accessed 20 April 2013) (Blom et al., 2009). The respective sizes of the core genomes were calculated using EDGAR, where shared orthologous genes were defined as those that shared ≥70% identity and that were aligned over at least 70% of the length of the query sequence.

The genome annotation data, including the COG classifications of the genes, for the isolates included was downloaded from the database of the National Centre for Biotechnology Information (NCBI; http://0-www.ncbi.nlm.nih.gov.innopac.up.ac.za/genome). The annotation data was then mapped to the genes present in the core as obtained from EDGAR. Visualization of these results was also performed using TreeMap 2.7.5 from Macrofocus, available at www.treemap.com.

Calculation of average amino acid identities for individual genera

The AAI of the genera were calculated to determine whether each of the genera that were compared could be considered as a well-defined genus based on the proposed AAI cut-off value. These values were determined by using the amino acid sequences, obtained from EDGAR, for 3 species of each genus. The species were selected based on their phylogenetic diversity as indicated in EDGAR, where the AAI analyses included the most closely related species and the most distantly related species (See Tables 1). The AAI calculations utilized standard parameters (Konstantinidis and Tiedje, 2005a; Konstantinidis and Tiedje, 2005b), where the average identities for the protein sequences aligned over more than 70% of the length of the query sequence, with an identity of at least 30%.

Determination of the Pantoea *core genome*

The core genome of the genus *Pantoea* was determined using all 18 of the available genome sequences for isolates and species in this genus (Table 2). These genome sequences thus spanned the known diversity of the genus comprehensively (i.e. the core genome determinations using EDGAR database utilized only six genomes representing four species of *Pantoea*). Within the respective genome sequences, protein-coding sequences were identified using the Bacterial gene finding tool (FGENESB) (Solovyev and Salamov, 2011), available from the Softberry platform (http://linux1.softberry.com/berry.phtml). For this purpose, the basic gene set of the well-annotated *P. ananatis* AJ13355 genome was used.

The core genome of *P. ananatis,* for which the genomes of 8 isolates were available, was first determined. This was done by comparing all the protein sequences of genes from all the *P. ananatis* isolates with one another using blastp (Altschul et al., 1997) from Blast 2.2.25 + as an accessory application of BioEdit 7.0.9. Genes were considered as core gene homologues if they shared ≥30% identity and aligned over at least 70% of the length of the query sequence (Konstantinidis and Tiedje, 2005a). The protein sequences for genes present in all the *P. ananatis* strains were then compared to the other genomes available using blastp and all the protein sequences for genes present in the compared genomes were extracted from the *P. ananatis* core file. This was repeated until a file containing only the protein sequences for genes present in all of the genomes included, was obtained. The final core set of genes obtained for *Pantoea* was classified in terms of their COGs by comparison to the COG database by using BLAST 2.2.25+ as an accessory application of BioEdit 7.0.9.

For the comparison of gene content to the *T. morbirosei* genome, the same parameters were used as in the *Pantoea* core calculation. Thus at least 30% identity was required over at least 70% alignment over the length of the query sequence. This was done to determine if the core genome of *Pantoea* remained more cohesive than the shared gene content of isolates from different genera. These more relaxed parameters were also applied as the isolates used ranged from closely related to more distantly related organisms, compared to the genomes used in EDGAR.

Results

Core genome sizes based on EDGAR

The core genomes of *Pantoea, Erwinia, Escherichia*, *Ralstonia, Burkholderia, Shewanella, Rhizobium, Listeria* and *Lactobacillus* were determined from the genomes available on EDGAR. The core genomes for these genera varied considerably and consisted of 2904, 2295, 1435, 2074, 1096, 1769, 2389, 1973 and 169 genes, respectively. The core genomes were divided into general cellular processes, namely ‘Cellular processing and signalling’, ‘Information storage and processing’, ‘Metabolism’, ‘Poorly Characterized’ and genes with no homologues in the COG database (Table 3).

For *Pantoea*, *Ralstonia* and *Erwinia*, their respective core sets of genes reduced by almost 50% (Table 4) with the addition of an isolate of their sister genera. With the addition of the *T. morbirosei* sequence to the *Pantoea* core genome, the shared genome content reduced from 2907 genes to 1506 genes. The shared genome content between the *Erwinia* core genome (2295) and the *P. ananatis* isolate was only 1256, causing a reduction of more than 1000 genes. With the addition of the *C.*

taiwanensis sequence to the *Ralstonia* core genome, a reduction in shared gene content of about 800 genes was observed (2074 to 1230).

Comparison of the core genomes

a. Average amino acid identities

The range of AAI values between species was calculated by determining the AAI value between the two phylogenetically most distant and the two most closely related species with genomes available in EDGAR (Table 1). The values for the individual genera ranged from 56.42-72.94% for *Lactobacillus* to as high as 85.56-97.2% for *Escherichia* (Table 5).

b. COG classifications

Based on the EDGAR calculations, the core genome for *Pantoea* was the largest of all genera compared. Grouping of the sequences into general cellular processes (Appendix A, Table1) showed the highest number of genes to be involved in metabolism. Six hundred and thirty-five genes were involved in ‘Cellular processes and signalling’, 534 genes were involved in ‘Information storage and processing’, 1248 genes were involved in ‘Metabolism’, 564 genes were ‘Poorly characterized’ and 221 genes had no homologues in the COG database (Figure 2, Table 3). Some genes were associated with multiple functional categories and were thus involved in more than one cellular process. An example of this is PAJ_0018 (*Pantoea ananatis* AJ13355 gene), annotated as an orthologue of *IspH*, which is involved in ‘lipid transport and metabolism’, as well as ‘cell wall/ membrane/ envelope biogenesis’, as it is associated with penicillin tolerance. This gene was thus present under both ‘cellular processes and signalling’ and ‘metabolism’.

The fraction of core genes associated with general cellular processes are quite similar for all genera, independent of the number of genes present in the specific core genome (Figure 3; Appendix A, Figure 1 - 8). The fraction of genes associated with each functional category also does not differ greatly between the compared genera (Figure 5). By comparing the core genomes of the nine different genera (Figure 3), it was observed that the highest fraction of genes in the core was generally involved in ‘metabolism’ (32.00% - 46.9%), followed by ‘cellular processes and signalling’ (16.7% - 25.2%), then by ‘information storage and processing’ (16.9% - 20.3%), followed by ‘poorly characterized genes’ (13.3% - 20.5%) and finally by genes with no associated COG functions (1.9% - 9.5%). The only exception was the core genome of *Lactobacillus* where the highest proportion of genes was linked to ‘information storage and processing’ (37.6%), followed by ‘metabolism’ (27.5%), and then by ‘cellular processes and signalling’ genes (20.2%) and ‘poorly characterized’ genes (14.6%).

In all functional categories, except six of the categories (‘Cell motility’ (N), ‘Intracellular trafficking, secretion and vesicular transport’ (U), ‘Chromatin structure and dynamics’ (B), ‘RNA processing and modification’ (A), ‘Lipid transport and metabolism’ (I) and ‘Secondary metabolite biosynthesis, transport and catabolism’ (Q)), *Pantoea* had the highest number of genes (Figure 4 and Table 3). A high number of genes in *Pantoea* was associated with metabolism, especially in the functional categories ‘Amino acid transport and metabolism’ (E), ‘Carbohydrate transport and metabolism’ (G) and ‘Inorganic ion transport and metabolism’ (P). A high number of genes were also involved in ‘Cell wall/ membrane/ envelope biogenesis’ (M) and ‘Transcription’ (K). There was also a relatively high number of genes (308) that only had general function predictions, as they were not yet fully characterized ('General function prediction only' (R)).

When focussing on the fraction of genes that were present in each functional category in relation to the size of the core (Figure 5), it appeared that the ratio of genes in each functional category remained relatively constant in all the genera investigated. The exception to this was *Lactobacillus* with its highly reduced core. Functional categories with a high proportion of genes for *Lactobacillus* were those responsible for cell wall and membrane biosynthesis (M), replication (L) and translation (J), which are essential for survival. This trend was also observed to a lesser extent for *Burkholderia*, where the fraction of genes in the category 'Translation, ribosomal structure and biogenesis' (J) is much higher than for the other functional categories. Only minor differences could be observed between the other genera, such as in *Escherichia* and *Rhizobium* that had a high fraction of genes involved in 'Amino acid transport and metabolism' (E) and *Ralstonia* that had a higher fraction of genes involved in 'Lipid transport and metabolism' (I) compared to the other genera.

Core genome size of Pantoea based on 18 WGSs

To ensure that the massive reduction observed were not caused by the limited number of genomes used or the parameters employed, the core genome of *Pantoea* was manually calculated using more genomes and thereafter compared to an isolate of the sister-genus *Tatumella* using more relaxed parameters. This core genome for *Pantoea,* based on 18 genome sequences, consisted of 2491 genes. The COG groupings and other results obtained for the manually calculated core genome of *Pantoea* correlated well with the results for the core genome obtained from EDGAR.

Grouping of these genes according to general cellular processes still showed the highest number of genes to be involved in metabolism. Five hundred and fifty-nine genes were involved in ‘Cellular processing and signalling’, 465 genes were involved in ‘Information storage and processing’, 1115 genes were involved in ‘Metabolism’, 516 genes were ‘Poorly characterized’ and 184 genes had no associated COG description in the COG database (Table 6).

As each species of *Pantoea* was added to the list of core genes, the number of genes to be removed from the list decreased (Figure 6). However, when the genome sequence of *T. morbirosei* was added to the comparison as a representative of the sister-genus *Tatumella*, a sharp reduction in shared genes were again observed. The dataset containing the *T. morbirosei* whole genome sequence in addition to those of the 18 *Pantoea* isolates contained only 1275 shared genes as opposed to the 2491 shared within the genus *Pantoea*.

In terms of the COG classifications of the *Pantoea* and the *Pantoea* + *Tatumella* cores, the largest reduction (of more than 60%) was observed amongst the proteins that are poorly characterized (516 proteins to 196 proteins) or had no COG association (184 proteins to 60 proteins), suggesting that many of these uncharacterized genes are unique to *Pantoea* isolates. A reduction of approximately 50% was observed for proteins involved in cellular processing and signalling (559 proteins to 264 proteins) and metabolism (1115 proteins to 515 proteins). The reduction of proteins involved in information storage and processing (465 proteins to 289 proteins) was the smallest with a loss of about 40% of the genes (Table 6).

Discussion

The results of this study suggest a high level of genetic and functional cohesiveness in the core genome of *Pantoea*. This was evident from the considerable size of the set of genes shared by all the examined isolates. The *Pantoea* core genome is larger than those observed for the other genera evaluated, and is made up of genes typically associated with a core genome as well as a large number of genes involved in metabolism. The cohesiveness of the *Pantoea* core was also apparent as a large fraction of these genes were not shared with its sister-taxon, *Tatumella*. In fact, a similar reduction in core genome size was also observed when the core gene sets *Erwinia* and *Ralstonia* were compared with close of their sister genera. Overall, these observations thus support the hypothesis that the core genome of a bacterial genus could be used to delineate the boundaries of the genus.

The core genome of *Pantoea*, as calculated in EDGAR, consisted of more than 50% of the genes present in an average *Pantoea* genome. This is relatively high as most genera used in comparative studies were reported to have a much smaller fraction of conserved genes (Canchaya et al., 2006; Lefébure and Stanhope, 2007; Prasanna and Mehra, 2013). An example of this is that of *Streptococcus* where a core genome of 26 genomes consisted of approximately 600 genes (Lefébure and Stanhope, 2007), less than a third of the average *Streptococcus* genome size. In extreme cases where a very large accessory genome is present, such as in *E. coli*, even the species core can be as low as 20% of the average genome size (Lukjancenko et al., 2012).

The size and accuracy of a determined core genome are closely linked to the number of isolates investigated. As the number of sequenced isolates increase, the diversity within a group will become more apparent, and more genes will be grouped as part of the accessory genome (Lukjancenko et al., 2012). During the manual calculation of the *Pantoea* core genome, an initial sharp drop in gene content was observed after which it gradually declined and eventually plateaued. This was expected as it has been observed that comparisons of more genomes of different species within the same genus causes the size of the core genome to decrease (Lukjancenko et al., 2012) until most genes present in the core dataset is present in all new isolates added to the dataset.

Closely related organisms will be more similar in gene content and genomic signatures (Kunin et al., 2005), and a large core genome is expected. It is also expected that the size of the core genome will decrease slowly between closely related isolates with more drastic reductions when distantly related isolates are added to the comparison (Brown et al., 2001). In the current study, such large reductions in shared gene content between isolates of different genera were noted. This observation was not dependent on the number of genomes used or the strict parameters used for the determination of the core genomes, as this effect was not only observed for the core genomes of *Pantoea, Erwinia* and *Ralstonia* determined by EDGAR but also for the manually calculated *Pantoea* core genome. The manually calculated core was based on more genomes and more relaxed homology parameters. With the addition of the genome of *T. morbirosei*, the manually calculated *Pantoea* core genome set was also reduced by more than 1000 genes. Species in the genus *Tatumella* used to be classified as *Pantoea* species, and based on the massive reduction on the core genome and the diversity in core genome content it appears that the separation of these genera was well-grounded.

The massive intergeneric reduction in the core genomes, may also explain the small core genome sizes observed for certain genera. The core genome determined in this study for *Lactobacillus* only consisted of 169 genes. This is much smaller than expected when compared to the other genera. The small number of shared genes may be due to the diverse nature of the species currently defined as belonging to the genus *Lactobacillus*. The core genome data confirmed that these species are not only phenotypically very diverse, but also genotypically (Felis and Dellaglio, 2007). In this example the small core genome for the genus may be the result of the inclusion of several potential genera with their corresponding diverse core genomes, within this single higher taxon.

The lowest AAI values calculated for genome pairs belonging to the same genus, were mostly above the minimum limit of 70%, the proposed genus cut-off value (Konstantinidis and Tiedje, 2007). The only exceptions in the current study were noted for *Rhizobium* and *Lactobacillus* species. In the case of the genus *Rhizobium*, *Agrobacterium* and *Rhizobium* was analysed as belonging to the same genus, as *Agrobacterium* was previously transferred to the genus *Rhizobium* based on their monophyletic grouping (Young et al., 2001). However, the value between *Rh. leguminosarum* and *A. radiobacter* was below 70%. Based on this AAI value, revision of the taxonomy of the genus *Rhizobium* (suggested by Mousavi et al., 2014) is supported. The low AAI values observed for *Lactobacillus* species (56.42% - 72.94%), together with the extremely small core genome for the genus, suggests that the genus is also too diverse as it is currently defined, and requires taxonomic revision (Canchaya et al., 2006; Makarova et al., 2006). The 70% genus cut-off value suggested by Konstantinidis and Tiedje (2007) appears to support the existence of the majority of the genera analysed in this study. It may, however, be necessary to look at the core genomes of the genera individually and not just to apply a generic cut-off value for genus descriptions as exceptions are sure to be found.

Based on all the available data, it appears that intra-generic cohesiveness could also be linked to the biology. A high number of metabolic core genes in *Pantoea,* reflected the shared metabolic diversity observed for this genus. Although the lifestyles of species in this genus, which ranges from opportunistic pathogens to environmental species, varied (De Champs et al., 2000; De Baere et al., 2004; Medrano and Bell, 2007; Brady et al., 2008; Cataño et al., 2012; De Maayer et al., 2012a; 2012b) several metabolic functional classes were conserved among all members of *Pantoea*, irrespective of their lifestyles. These findings suggest that, despite the accessory genome's involvement in niche specialization and adaptation, the core genome of *Pantoea* encode all the necessary machinery to allow its survival in a range of diverse conditions. Conservation of this high number of genes, usually not considered as essential for survival, could also be linked to the ability of these species to survive in the environment, when they are not associated with the usual host(s) (Gavini et al., 1989; Brady et al., 2008).

A significant fraction of the *Pantoea* core genome appears to encode products with no known function or products that only have general functional predictions. As similar genes are captured in the COG database, these genes are thus conserved in a number of species and are not unique to *Pantoea*. Similar studies have also reported a relatively large number of core genes that have only general function predictions and have not yet been characterized (Canchaya et al., 2006; Bennett et al., 2010; Prasanna and Mehra, 2013). This demonstrates the need for functional studies to determine the function and role of these genes in cellular processes.

Some of the *Pantoea* core genes had no matches in the COG database. This may either be due to paralogy, as the COG database only contains gene families with no known paralogues in a single lineage (Tatusov et al., 2000), or because these genes are unique to *Pantoea* or new to science. Further investigation is, however, required to determine which scenario is more likely. There are also a number of genes in the core genomes of other genera that fall into this class (Canchaya et al., 2006; Bennett et al., 2010; Prasanna and Mehra, 2013), suggesting that this is not unusual.

The number of genes decreased in all categories as the core genome decreased in size. However, fewer genes involved in 'Information storage and processing' were eliminated from the core when compared to the other categories of cellular processes. In the highly reduced core genomes like *Lactobacillus*, and to a lesser extent *Burkholderia*, there is thus a bias towards more informational genes retained in the core. For the core genomes of the other genera the relative ratio of genes involved in each cellular process remained fairly constant as they still possessed larger core genomes. The lifestyle of individual species and diversity observed within a genus thus impact on the size of the core genome, but it appears that a reduction of genes in all functional categories occur as the core decreases. This gradual reduction only occurs up to a certain point, after which the decrease in genes involved in 'Information storage and processing' is decelerated, and these indispensable genes are retained. Although many attributes, like metabolism of certain carbohydrates and lipids, may not be shared by all members of the genus, these informational genes have to be present for basic survival.

The COG classes that remained conserved between the manually calculated core genome of *Pantoea* and the *Pantoea* + *Tatumella* core genomes supported the idea that the addition of too diverse taxa causes a massive decrease in the core genome size through a reduction in proteins from all cellular processes. It does, however, appear that even at this scale (between genera), the reduction in the uncharacterized and unknown genes are much larger than for the more essential genes. It also supports the notion that a slower reduction in information storage and processing proteins occurs as these genes are expected to remain more conserved in divergent groups of organisms as they are essential for survival.

The findings presented in this study indicate that core genome data can be used to determine boundaries between genera. Apart from representing a valuable taxonomic tool, the use of this information can also provide an indication of biological characteristics that are conserved within genera and potentially reveal metabolic, ecological and general biological properties unique to the taxa in question. This tool may also be able to assist in delineating higher taxa. The use of core genome data for higher level delineation of taxa will undoubtedly become more feasible as WGS data, representative of the phylogenetic diversity of bacteria, progressively increases.

References

ADAMS, A. S., JORDAN, M. S., ADAMS, S. M., SUEN, G., GOODWIN, L. A., DAVENPORT, K. W., CURRIE, C. R. & RAFFA, K. F. (2011) Cellulose-degrading bacteria associated with the invasive woodwasp *Sirex noctilio*. *ISME Journal,* 5, 1323-1331.

ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25, 3389-3402.

AMADOU, C., PASCAL, G. R., MANGENOT, S., GLEW, M., BONTEMPS, C., CAPELA, D., CARRÉRE, S. B., CRUVEILLER, S. P., DOSSAT, C., LAJUS, A. L., MARCHETTI, M., POINSOT, V. R. N., ROUY, Z., SERVIN, B., SAAD, M., SCHENOWITZ, C., BARBE, V. R., BATUT, J., MÉDIGUE, C. & MASSON-BOIVIN, C. (2008) Genome sequence of the β-rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome Research,* 18, 1472-1483.

BENNETT, J., BENTLEY, S., VERNIKOS, G., QUAIL, M., CHEREVACH, I., WHITE, B., PARKHILL, J. & MAIDEN, M. (2010) Independent evolution of the core and accessory gene sets in the genus *Neisseria*: insights gained from the genome of *Neisseria lactamica* isolate 020-06. *BMC Genomics,* 11, 652.

BLOM, J., ALBAUM, S., DOPPMEIER, D., PUHLER, A., VORHOLTER, F.-J., ZAKRZEWSKI, M. & GOESMANN, A. (2009) EDGAR: A software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics,* 10, 154.

BRADY, C., CLEENWERCK, I., VENTER, S., VANCANNEYT, M., SWINGS, J. & COUTINHO, T. (2008) Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Systematic and Applied Microbiology,* 31, 447-460.

BRADY, C., VENTER, S., CLEENWERCK, I., VANCANNEYT, M., SWINGS, J. & COUTINHO, T. (2007) A FAFLP system for the improved identification of plant-pathogenic and plant-associated species of the genus Pantoea. *Systematic and Applied Microbiology,* 30, 413-417.

BRADY, C. L., CLEENWERCK, I., VENTER, S. N., ENGELBEEN, K., DE VOS, P. & COUTINHO, T. A. (2010a) Emended description of the genus *Pantoea*, description of four species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov., and transfer of *Pectobacterium cypripedii* (Hori 1911) Brenner et al. 1973 emend. Hauben et al. 1998 to the genus as *Pantoea cypripedii* comb. nov. *International Journal of Systematic and Evolutionary Microbiology,* 60, 2430-2440.

BRADY, C. L., VENTER, S. N., CLEENWERCK, I., ENGELBEEN, K., VANCANNEYT, M., SWINGS, J. & COUTINHO, T. A. (2009) *Pantoea vagans* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyi* sp. nov. and *Pantoea anthophila* sp. nov. *International Journal of Systematic and Evolutionary Microbiology,* 59, 2339-2345.

BRADY, C. L., VENTER, S. N., CLEENWERCK, I., VANDEMEULEBROECKE, K., DE VOS, P. & COUTINHO, T. A. (2010b) Transfer of *Pantoea citrea*, *Pantoea punctata* and *Pantoea terrea* to the genus *Tatumella* emend. as *Tatumella citrea* comb. nov., *Tatumella punctata* comb. nov. and *Tatumella terrea* comb. nov. and description of *Tatumella morbirosei* sp. nov. *International Journal of Systematic and Evolutionary Microbiology,* 60, 484-494.

BROWN, J. R., DOUADY, C. J., ITALIA, M. J., MARSHALL, W. E. & STANHOPE, M. J. (2001) Universal trees based on large combined protein sequence data sets. *Nat Genet,* 28, 281-285.

CANCHAYA, C., CLAESSON, M. J., FITZGERALD, G. F., VAN SINDEREN, D. & O'TOOLE, P. W. (2006) Diversity of the genus *Lactobacillus* revealed by comparative genomics of five species. *Microbiology,* 152, 3185-3196.

CATAÑO, J. C., ECHEVERRI, L. M. & SZELA, C. (2012) Bacterial contamination of clothes and environmental items in a third-level hospital in Colombia. *Interdisciplinary Perspectives on Infectious Diseases,* 2012.

CHOI, O., LIM, J. Y., SEO, Y.-S., HWANG, I. & KIM, J. (2012) Complete genome sequence of the rice pathogen *Pantoea ananatis* strain PA13. *Journal of Bacteriology,* 194**,** 531.

CICCARELLI, F. D., DOERKS, T., VON MERING, C., CREEVEY, C. J., SNEL, B. & BORK, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science,* 311, 1283-1287.

COENYE, T., GEVERS, D., VAN DE PEER, Y., VANDAMME, P. & SWINGS, J. (2005) Towards a prokaryotic genomic taxonomy. *Federation of European Microbiological Societies Microbiology Reviews,* 29, 147-167.

COHAN, F. M. (2001) Bacterial species and speciation. *Systematic Biology,* 50, 513 - 524.

COMAS, I. A., MOYA, A. S. & GONZÃLEZ-CANDELAS, F. (2007) From phylogenetics to phylogenomics: The evolutionary relationships of insect endosymbiotic γ-Proteobacteria as a test case. *Systematic Biology,* 56, 1-16.

DAUBIN, V., GOUY, M. & PERRIÈRE, G. (2002) A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Research,* 12, 1080-1090.

DE BAERE, T., VERHELST, R., LABIT, C., VERSCHRAEGEN, G., WAUTERS, G., CLAEYS, G. & VANEECHOUTTE, M. (2004) Bacteremic infection with *Pantoea ananatis*. *Journal of Clinical Microbiology,* 42, 4393-4395.

DE CHAMPS, C., LE SEAUX, S., DUBOST, J. J., BOISGARD, S., SAUVEZIE, B. & SIROT, J. (2000) Isolation of *Pantoea agglomerans* in two cases of septic monoarthritis after plant thorn and wood sliver injuries. *Journal of Clinical Microbiology,* 38, 460-461.

DE MAAYER, P., CHAN, W.-Y., BLOM, J., VENTER, S., DUFFY, B., SMITS, T. & COUTINHO, T. (2012a) The large universal *Pantoea* plasmid LPP-1 plays a major role in biological and ecological diversification. *BMC Genomics,* 13, 625.

DE MAAYER, P., CHAN, W. Y., REZZONICO, F., BÜHLMANN, A., VENTER, S. N., BLOM, J., GOESMANN, A., FREY, J. E., SMITS, T. H. M., DUFFY, B. & COUTINHO, T. A. (2012b) Complete genome sequence of clinical isolate *Pantoea ananatis* LMG 5342. *Journal of Bacteriology,* 194, 1615-1616.

DE MAAYER, P., CHAN, W. Y., VENTER, S. N., TOTH, I. K., BIRCH, P. R. J., JOUBERT, F. & COUTINHO, T. A. (2010) Genome Sequence of *Pantoea ananatis* LMG20103, the causative agent of Eucalyptus blight and dieback. *Journal of Bacteriology,* 192, 2936-2937.

DOOLITTLE, F. W. & PAPKE, T. R. (2006) Genomics and the bacterial species problem. *Genome Biology,* 7, 116.

EDWARDS, S. V., LIU, L. & PEARL, D. K. (2007) High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences,* 104, 5936-5941.

FELIS, G. E. & DELLAGLIO, F. (2007) Taxonomy of lactobacilli and bifidobacteria. *Current issues in intestinal microbiology,* 8, 44.

GAO, B. & GUPTA, R. S. (2012) Microbial systematics in the post-genomics era. *Antonie Van Leeuwenhoek,* 101, 45-54.

GAVINI, F., MERGAERT, J., BEJI, A., MIELCAREK, C., IZARD, D., KERSTERS, K. & DE LEY, J. (1989) Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and Description of *Pantoea dispersa* sp. nov. *International Journal of Systematic Bacteriology,* 39, 337-345.

GEVERS, D., COHAN, F. M., LAWRENCE, J. G., SPRATT, B. G., COENYE, T., FEIL, E. J., STACKEBRANDT, E., VAN DE PEER, Y., VANDAMME, P., THOMPSON, F. L. & SWINGS, J. (2005) Re-evaluating prokaryotic species. *Nature Reviews,* 3, 733-739.

GORIS, J., KONSTANTINIDIS, K. T., KLAPPENBACH, J. A., COENYE, T., VANDAMME, P. & TIEDJE, J. M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology,* 57, 81-91.

HARA, Y., KADOTANI, N., IZUI, H., KATASHKINA, J., KUVAEVA, T., ANDREEVA, I., GOLUBEVA, L., MALKO, D., MAKEEV, V., MASHKO, S. & KOZLOV, Y. (2012) The complete genome sequence of Pantoea ananatis AJ13355, an organism with great biotechnological potential. *Applied Microbiology and Biotechnology,* 93**,** 331-341.

HACKER, J. R. & CARNIEL, E. (2001) Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO reports,* 2, 376-381.

HACKER, J. R. H., DOBRINDT, U. & KURTH, R. (2012) *Genome plasticity and infectious diseases*, ASM Press, Washington DC, USA.

KIM, H. J., LEE, J. H., KANG, B. R., RONG, X., MCSPADDEN GARDENER, B. B., JI, H. J., PARK, C.-S. & KIM, Y. C. (2012) Draft genome sequence of *Pantoea ananatis* B1-9, a nonpathogenic plant growth-promoting bacterium. *Journal of Bacteriology,* 194, 729.

KLENK, H. P. & GÖKER, M. (2010) En route to a genome-based classification of Archaea and Bacteria? *Systematic and Applied Microbiology,* 33, 175-182.

KOBAYASHI, H., KAWASAKI, K., TAKEISHI, K. & NODA, H. (2011) Symbiont of the stink bug *Plautia stali* synthesizes rough-type lipopolysaccharide. *Microbiological Research,* 167, 48-54.

KONSTANTINIDIS, K. T., RAMETTE, A. & TIEDJE, J. M. (2006) The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 361, 1929-1940.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. (2005a) Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences,* 102, 2567-2572.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. (2005b) Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology,* 187, 6258-6264.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. (2006) Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Applied and Environmental Microbiology,* 72, 7286-7293.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. (2007) Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Current Opinion in Microbiology,* 10, 504-509.

KUNIN, V., AHREN, D., GOLDOVSKY, L., JANSSEN, P. & OUZOUNIS, C. A. (2005) Measuring genome conservation across taxa: divided strains and united kingdoms. *Nucleic Acids Research,* 33, 616-621.

LAN, R. & REEVES, P. R. (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends in Microbiology,* 8, 396-401.

LEFÉBURE, T. & STANHOPE, M. J. (2007) Evolution of the core and pan-genome of Streptococcus: positive selection, recombination, and genome composition. *Genome Biology,* 8, R71.

LOPER, J. E., HASSAN, K. A., MAVRODI, D. V., DAVIS, E. W., II, LIM, C. K., SHAFFER, B. T., ELBOURNE, L. D. H., STOCKWELL, V. O., HARTNEY, S. L., BREAKWELL, K., HENKELS, M. D., TETU, S. G., RANGEL, L. I., KIDARSA, T. A., WILSON, N. L., VAN DE MORTEL, J. E., SONG, C., BLUMHAGEN, R., RADUNE, D., HOSTETLER, J. B., BRINKAC, L. M., DURKIN, A. S., KLUEPFEL, D. A., WECHTER, W. P., ANDERSON, A. J., KIM, Y. C., PIERSON, L. S., III, PIERSON, E. A., LINDOW, S. E., KOBAYASHI, D. Y., RAAIJMAKERS, J. M., WELLER, D. M., THOMASHOW, L. S., ALLEN, A. E. & PAULSEN, I. T. (2012) Comparative Genomics of plant-associated *Pseudomonas* spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genetics,* 8, e1002784.

LUKJANCENKO, O., WASSENAAR, T. & USSERY, D. (2012) Comparison of 61 sequenced *Escherichia coli* genomes. *Microbial Ecology,* 60, 708-720.

MAKAROVA, K., SLESAREV, A., WOLF, Y., SOROKIN, A., MIRKIN, B., KOONIN, E., PAVLOV, A., PAVLOVA, N., KARAMYCHEV, V., POLOUCHINE, N., SHAKHOVA, V., GRIGORIEV, I., LOU, Y., ROHKSAR, D., LUCAS, S., HUANG, K., GOODSTEIN, D. M., HAWKINS, T., PLENGVIDHYA, V., WELKER, D., HUGHES, J., GOH, Y., BENSON, A., BALDWIN, K., LEE, J. H., DÍAZ-MUÑIZ, I., DOSTI, B., SMEIANOV, V., WECHTER, W., BARABOTE, R., LORCA, G., ALTERMANN, E., BARRANGOU, R., GANESAN, B., XIE, Y., RAWSTHORNE, H., TAMIR, D., PARKER, C., BREIDT, F., BROADBENT, J., HUTKINS, R., O'SULLIVAN, D., STEELE, J., UNLU, G., SAIER, M., KLAENHAMMER, T., RICHARDSON, P., KOZYAVKIN, S., WEIMER, B. & MILLS, D. (2006) Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences,* 103, 15611-15616.

MANN, R. A., SMITS, T. H. M., BÜHLMANN, A., BLOM, J., GOESMANN, A., FREY, J. E., PLUMMER, K. M., BEER, S. V., LUCK, J., DUFFY, B. & RODONI, B. (2013) Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. *PLoS ONE,* 8, e55644.

MEDRANO, E. G. & BELL, A. A. (2007) Role of *Pantoea agglomerans* in opportunistic bacterial seed and boll rot of cotton (*Gossypium hirsutum*) grown in the field. *Journal of Applied Microbiology,* 102, 134-143.

MONTAÑEZ, A., BLANCO, A. R., BARLOCCO, C., BERACOCHEA, M. & SICARDI, M. (2012) Characterization of cultivable putative endophytic plant growth promoting. *Applied Soil Ecology,* 58, 7.

MOUSAVI, S. A., ÖSTERMAN, J., WAHLBERG, N., NESME, X., LAVIRE, C., VIAL, L., PAULIN, L., DE LAJUDIE, P. & LINDSTRÖM, K. (2014) Phylogeny of the *Rhizobium-Allorhizobium-Agrobacterium* clade supports the delineation of *Neorhizobium* gen. nov. *Systematic and Applied Microbiology*.

PRASANNA, A. N. & MEHRA, S. (2013) Comparative phylogenomics of pathogenic and non-pathogenic *Mycobacterium*. *PLoS ONE,* 8, e71248.

RICHTER, M. & ROSSELLÓ-MÓRA, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences,* 106, 19126-19131.

SCHUBERT, R. H. (1968) The taxonomy and nomenclature of the genus *Aeromonas* Kluyver and van Niel 1936 Part III. Suggestions on the definition of the genus *Aeromonas* Kluyver and van Niel 1936. *International Journal of Systematic Bacteriology,* 18, 1-7.

SOLOVYEV, V. & SALAMOV, A. (2011) Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics and its applications in agriculture, biomedicine and environmental studies*, 61-78.

SUEN, G., SCOTT, J. J., AYLWARD, F. O., ADAMS, S. M., TRINGE, S. G., PINTO-TOMÃ¡S, A. N. A., FOSTER, C. E., PAULY, M., WEIMER, P. J., BARRY, K. W., GOODWIN, L. A., BOUFFARD, P., LI, L., OSTERBERGER, J., HARKINS, T. T., SLATER, S. C., DONOHUE, T. J. & CURRIE, C. R. (2010) An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genetics,* 6, e1001129.

TATUSOV, R. L., GALPERIN, M. Y., NATALE, D. A. & KOONIN, E. V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research,* 28, 33-36.

TETTELIN, H., RILEY, D., CATTUTO, C. & MEDINI, D. (2008) Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology,* 11, 472-477.

THOMPSON, C., VICENTE, A., SOUZA, R., VASCONCELOS, A., VESTH, T., ALVES, N., USSERY, D., IIDA, T. & THOMPSON, F. (2009) Genomic taxonomy of vibrios. *BMC Evolutionary Biology,* 9, 258.

THOMPSON, C. C., VIEIRA, N. M., VICENTE, A. C. P. & THOMPSON, F. L. (2011) Towards a genome based taxonomy of *Mycoplasmas*. *Infection, Genetics and Evolution,* 11, 1798-1804.

TINDALL, B. J., ROSSELLÓ-MORA, R., BUSSE, H. J., LUDWIG, W. & KÄMPFER, P. (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. *International Journal of Systematic and Evolutionary Microbiology,* 60, 249-266.

WHITMAN, W. B., GOODFELLOW, M., KÄMPFER, P., BUSSE, H. J., TRUJILLO, M. E., SUZUKI, K. & LUDWIG, W. (2012) *Bergey's Manual® of Systematic Bacteriology,* New York, Springer.

WINSLOW, C. E., BROADHURST, J., BUCHANAN, R. E., KRUMWIEDE JR, C., ROGERS, L. A. & SMITH, G. H. (1917) The families and genera of the bacteria: preliminary report of the committee of the society of american bacteriologists on characterization and classification of bacterial types. *Journal of Bacteriology,* 2, 505.

WINSLOW, C. E., BROADHURST, J., BUCHANAN, R. E., KRUMWIEDE JR, C., ROGERS, L. A. & SMITH, G. H. (1920) The families and genera of the bacteria: final report of the committee of the Society of American Bacteriologists on characterization and classification of bacterial types. *Journal of Bacteriology,* 5, 191.

YOUNG, J. M., KUYKENDALL, L. D., MARTÍNEZ-ROMERO, E., KERR, A. & SAWADA, H. (2001) A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. *International Journal of Systematic and Evolutionary Microbiology,* 51, 89-103.

Table 1 Genome sequences for the different genera investigated obtained from EGDAR*

Genus	# of genomes	Isolates							
Pantoea	6	*P. ananatis* LMG20103, LMG5342, AJ13355		*P. agglomerans* E325**	*P. vagans* C9-1 **	*P.* sp. At-9b**			
Erwinia	7	*Er. amylovora* ATC49946**, CFB1430	*Er. bilingiae* Eb661**	*Er. pyrifoliae* DSM12163, Ep 1-96	*Er. tasmaniensis* Et 1-99**	*Er.* sp. Ejp617			
Escherichia	32	*E. coli* 042**, ABU 83972, APEC 01, E24377A, ETEC H10407, K011FL (both sequences), O03:H2 strain 12009, O111:H strain 11128, O127:H6 strain E2348-69, O157:H7 strains EC4115, EDL93, Sakai, TW14359, O26:H11 strain 11368, O55:H7 strains CB9615, RM12579, O7:K1 strain CE10, O83:H1 strain NRG857C, S88, SE11, SE15, SMS-3-5, UM146, UMN026, UMNK88, UTI89, W (both sequences), Xuzhou 21						*E. fergusonii* ATCC35469**	*E. blattae* DSN4481**
Ralstonia	8	*Ra. eutropha* H16**, JMP134	*Ra. picketti* 12D**, 12J	*Ra. solanacearum* CFBP2957**, GMI1000, PSI07, Po82					
Burkholderia	30	*B. ambifaria* AMMD**, MC40-6	*B. cenocepacia* AU1054**, HI2424, J2315, MC0-3	*B. gladioli* BSR3	*B. glumae* BGR 1	*B. mallei* ATCC 2344, NCTC 10299, NCTC 10247, SAVP 1	*B. multivorans* ATCC 17616	*B. phymatum* STM 815	*B. phytofirmans* PsJN
		B. pseudomallei 1026b, 1106a, 1710b, 668, K96243	*B. rhizoxini* HKI454**	*B. thailandensis* E264	*B. vietnamensis* G4	*B. xenovorans* LB400	*B.* sp. 383, CCGE 1001, CCGE 1002, CCGE 1003, KJ006, YI23		
Shewanella	23	*S. baltica* BA175**, OS117, OS155, OS185, OS195, OS223, OS678	*S. oneidensis* MR-1	*S. amazonensis* SB2B	*S. denitrificans* OS217	*S. frigidimarina* NCIMB 400	*S. halifaxensis* HAW-EB4**	*S. loihica* PV-4	*S. pealeana* ATCC 700345**
		S. piezotolerans WP3	*S. putrefaciens* 200, CN-32	*S. sediminis* AW-EB3	*S. violacea* DSS12	*S. woodyi* ATCC51908	*S.* sp. ANA-3, MR-7, MR-4		

Table 1 – Continued

Genus	**# of genomes**	**Isolates**							
Rhizobium	10	*Agrobacterium radiobacter* K84**	*A.* sp. H13-3	*A. tumefaciens* C58	*A. vitis* S4	*Rh. etli* CFN 42**, CIAT652	*Rh. leguminosarum* WSM 1325**, WSM 2304 (bv. *trifolii),* 3841 (bv. *viviae*)	*Sinorhizobium fredii* NGR 234	
Listeria	17	*Li. innocua* Clip11262	*Li. ivanovii* ssp. *ivanovii* PAM55**	*Li. monocytogenes* 08-5578**, 07PF0776, 08-5923, 10403s, Clip81459, EGD-E, FSL R2-561, Finland 1998, HCC23, J016, L99, M7, F2365	*Li. seeligeri* SLCC3954**	*Li. welshimeri* SLCC5334			
Lactobacillus	35	*La. acidophilus* 30SC, NCFM	*La. amylovorus* GRL1118**, GRL1112	*La. brevis* ATCC367	*La. buchneri* NRRLB-3092	*La. casei* ATCC334, BD-II, BL23, LC2W, Zhang	*La. crispatus* ST1	*La. delbrueckii* ssp. *bulgaricus* ND02, 2038	*La. fermentum* CECT5716**, IFO3956
		La. gasseri ATCC33323	*La. helveticus* H10, DPC4571	*La. johnsonii* FI9785, NCC533	*La. kefiranofaciens* ZW3	*La. plantarum* JDM1, WCFS1, ST-III	*La. reuteri* DSM20016, JCM 1112**, SD2112	*La. rhamnosus* ATCC8530, GG, Lc705	*La. ruminis* ATCC27762
		La. sakei ssp. *sakei* 23K	*La. salivarius* CECT5713, UCC118	*La. sanfranciscensis* TMW11304					

* Efficient Database framework for comparative Genome Analyses using BLAST score Ratios (Blom et al., 2009); http://edgar.cebitec.uni-bielefeld.de/cgi-bin/edgar_login.cgi; accessed 20 April 2013;

**Genomes used to obtain protein-coding sequences for the calculation of Average Amino Acid Identity (AAI) values

Table 2 *Pantoea* genomes used for the manually calculated core genome of the genus.

Species	Lifestyle	Reference
P. agglomerans **E325**	Apple biocontrol agent	*1
P. agglomerans **MP2**	Termite symbiont	*2
P.(*agglomerans*) SL1-M5	Woodwasp symbiont	(Adams et al., 2011)
P. ananatis **AJ13355**	Saprophyte	(Hara et al., 2012)
P. ananatis **B1-9**	Onion growth-promoter	(Kim et al., 2012)
P.ananatis **BD442**	Maize pathogen	#
P. ananatis **LMG20103**	Eucalyptus pathogen	#(De Maayer et al., 2010)
P. ananatis **LMG2665 T**	Pineapple pathogen	#
P. ananatis **LMG5342**	Human pathogen	#(De Maayer et al., 2012)
P. ananatis **PA13**	Rice pathogen	(Choi et al., 2012)
P. ananatis **PA4**	Onion pathogen	#
P. (*dispersa*)	Stink bug symbiont	(Kobayashi et al., 2011)
P. (*eucalypti*) aB	Bark beetle symbiont	NCB (Unpublished): ASM17965v1
P. sp. At-9b	Leaf-cutter ant symbiont	(Suen et al., 2010)
P. stewartii **ssp.** ***indologenes*** **LMG2632 T**	Millet pathogen	#
P. stewartii **ssp.** ***stewartii*** **DC283**	Maize pathogen	*3
P. vagans **C9-1**	Apple biocontrol agent	*1
P. vagans **MP7**	Termite symbiont	*2

Genomes sequenced by the Department of Microbiology and Plant Pathology, University of Pretoria.

* Genome sequences obtained from collaborators ([1]Plant Protection Division, Agroscope Changins-Wädenswil, Switzerland; [2]Department of Biology, University of Copenhagen, Denmark; [3]Genome Evolution Laboratory, University of Wisconsin-Madison, USA).

Table 3 The number of core genes of each genus involved in the cellular processes and the subdivision of these genes into the COG functional categories.

General cellular processes	***Pantoea***	***Erwinia***	***Escherichia***	***Shewanella***	***Ralstonia***	***Burkholderia***	***Rhizobium***	***Listeria***	***Lactobacillus***
Cellular processes and signalling	**635**	**531**	**292**	**492**	**497**	**256**	**339**	**358**	**36**
[D] Cell cycle control, cell division, chromosome partitioning	30	27	21	27	24	21	20	29	3
[N] Cell motility	62	66	0	70	54	9	29	37	0
[M] Cell wall/ membrane/ envelope biogenesis	183	143	99	103	120	88	79	66	15
[V] Defence mechanisms	36	25	22	23	19	11	16	39	0
[U] Intracellular trafficking, secretion, and vesicular transport	64	68	18	78	80	33	39	39	3
[O] Post-translational modification, protein turnover, and chaperones	118	98	79	105	106	60	80	61	10
[T] Signal transduction mechanisms	142	105	53	85	95	34	77	87	5
Information storage and processing	**534**	**414**	**294**	**351**	**388**	**243**	**367**	**394**	**67**
[B] Chromatin structure and dynamics	0	0	0	0	1	1	1	0	0
[L] Replication, recombination and repair	123	105	77	102	95	63	81	95	23
[A] RNA processing and modification	1	1	1	1	2	1	0	0	0
[K] Transcription	228	153	91	100	140	64	131	162	13
[J] Translation, ribosomal structure and biogenesis	182	154	124	148	151	114	154	137	31
Metabolism	**1248**	**877**	**725**	**625**	**878**	**476**	**883**	**807**	**49**
[E] Amino acid transport and metabolism	309	240	197	149	200	104	255	184	6
[G] Carbohydrate transport and metabolism	259	138	108	53	93	48	140	165	15
[H] Coenzyme transport and metabolism	147	109	94	110	112	75	92	97	5
[C] Energy production and conversion	164	109	120	116	160	85	125	99	6
[P] Inorganic ion transport and metabolism	171	121	83	59	102	51	100	111	7
[I] Lipid transport and metabolism	74	54	43	68	108	47	65	52	2
[F] Nucleotide transport and metabolism	85	75	61	51	59	44	73	67	8
[Q] Secondary metabolites biosynthesis, transport, and catabolism	39	31	18	19	45	22	32	32	0
Poorly characterized	**564**	**394**	**206**	**308**	**415**	**177**	**365**	**439**	**26**
[S] Function unknown	249	191	84	150	177	72	145	219	5
[R] General function prediction only	315	203	122	158	239	105	221	220	21
No related COG	**221**	**233**	**29**	**178**	**86**	**43**	**50**	**146**	**0**

Table 4 The reduction in core genome size with the addition of a representative of the sister genus.

Genus	**Genus core from EDGAR**	**Addition of taxon from sister genus ***
Pantoea	2907	1506
Erwinia	2295	1256
Ralstonia	2074	1230

* *Pantoea* = *Tatumella morbirosei* LMG 2336 T (Department of Microbiology and Plant Pathology, University of Pretoria)

Erwinia = *Pantoea ananatis* LMG 20103 (De Maayer et al., 2010)

Ralstonia = *Cupriavidus taiwanensis* LMG 19424 T (Amadou et al., 2008)

Table 5 The genome architecture of the genera investigated

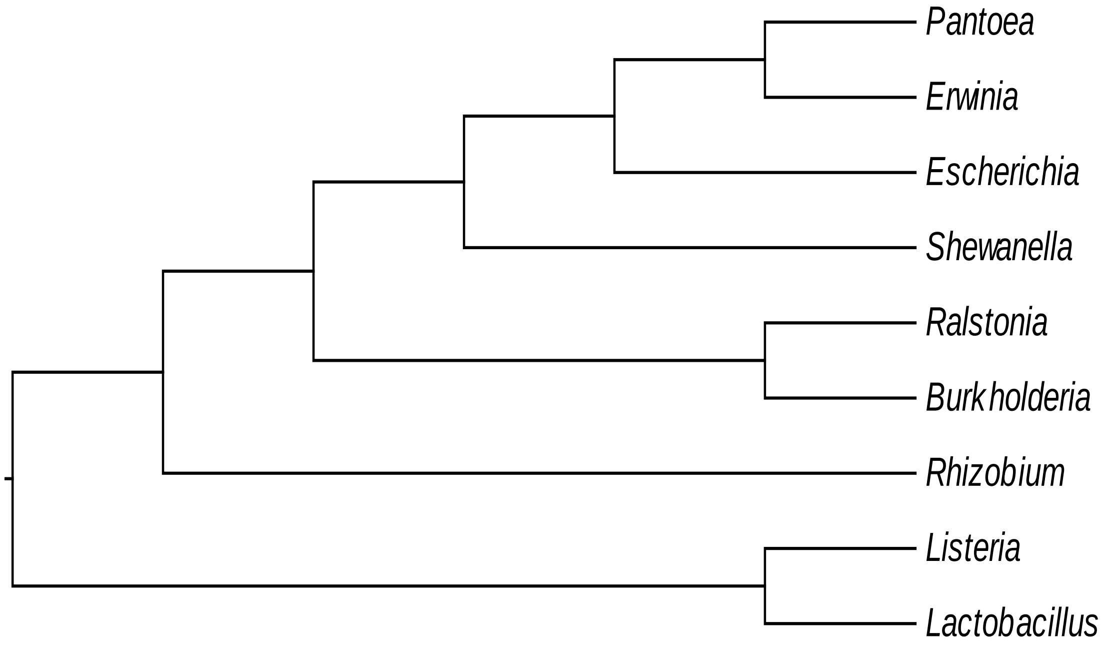
Genus	Genome size (bp) *	Number of genes*	AAI between species	Size of core genome**
Pantoea	4,580,000 - 6,310,000	4,000 - 6,000	77.90% - 92.06%	2907
Erwinia	3,800,000 - 5,370,000	3,700 - 5,029	82.51% - 91.20%	2295
Escherichia	4,030,000 - 5,590,000	3,800 - 5,500	85.56% - 97.2%	1435
Ralstonia	3,420,000 - 5,810,000	3,300 - 5,500	70.86% - 87.47%	2074
Burkholderia	6,150,000 - 7,880,000	5,700 - 7,900	70.66% - 94.58%	1096
Shewanella	4,700,000 - 5,940,000	4,000 - 5,100	72.68% - 90.46%	1769
Rhizobium	6,100,000 - 7,750,000	5,300 - 6,500	65.98% - 88.50%	2389
Listeria	2,900,000 - 3,140,000	2,900 - 3,200	90.42% - 93.79%	1973
Lactobacillus	1,700,000 - 2,600,000	1,700 - 2,800	56.42% - 72.94%	169

*Obtained from the National Centre for Biotechnology Information (NCBI; http://0-www.ncbi.nlm.nih.gov.innopac.up.ac.za/ genome)

** Obtained from EDGAR (Blom et al., 2009); http://edgar.cebitec.uni-bielefeld.de/cgi-bin/edgar_login.cgi; accessed 20 April 2013

Table 6 The reduction of genes in each general cellular process with the addition of a sister-taxon.

General cellular process	*Pantoea* core	*Pantoea* + *Tatumella* core
Cellular processing and signalling	559	264
Information storage and processing	465	289
Metabolism	1115	515
Poorly Characterized	516	196
No homologues in COG database	184	60



Average number of genes	AAI between species	Core genome size
5000	77.90% - 92.06%	2907
4350	82.51% - 91.20%	2295
4650	85.56% - 97.20%	1435
4550	72.68% - 90.46%	1769
4400	70.86% - 87.47%	2074
6800	70.66% - 94.58%	1096
5900	65.98% - 88.50%	2389
3050	90.42% - 93.79%	1973
2250	56.42% - 72.94%	169

Figure 1 The phylogenetic relationships of the genera used in the study as inferred from literature. The genome architecture of each genus (Table 5) is also visible next to the corresponding genera.

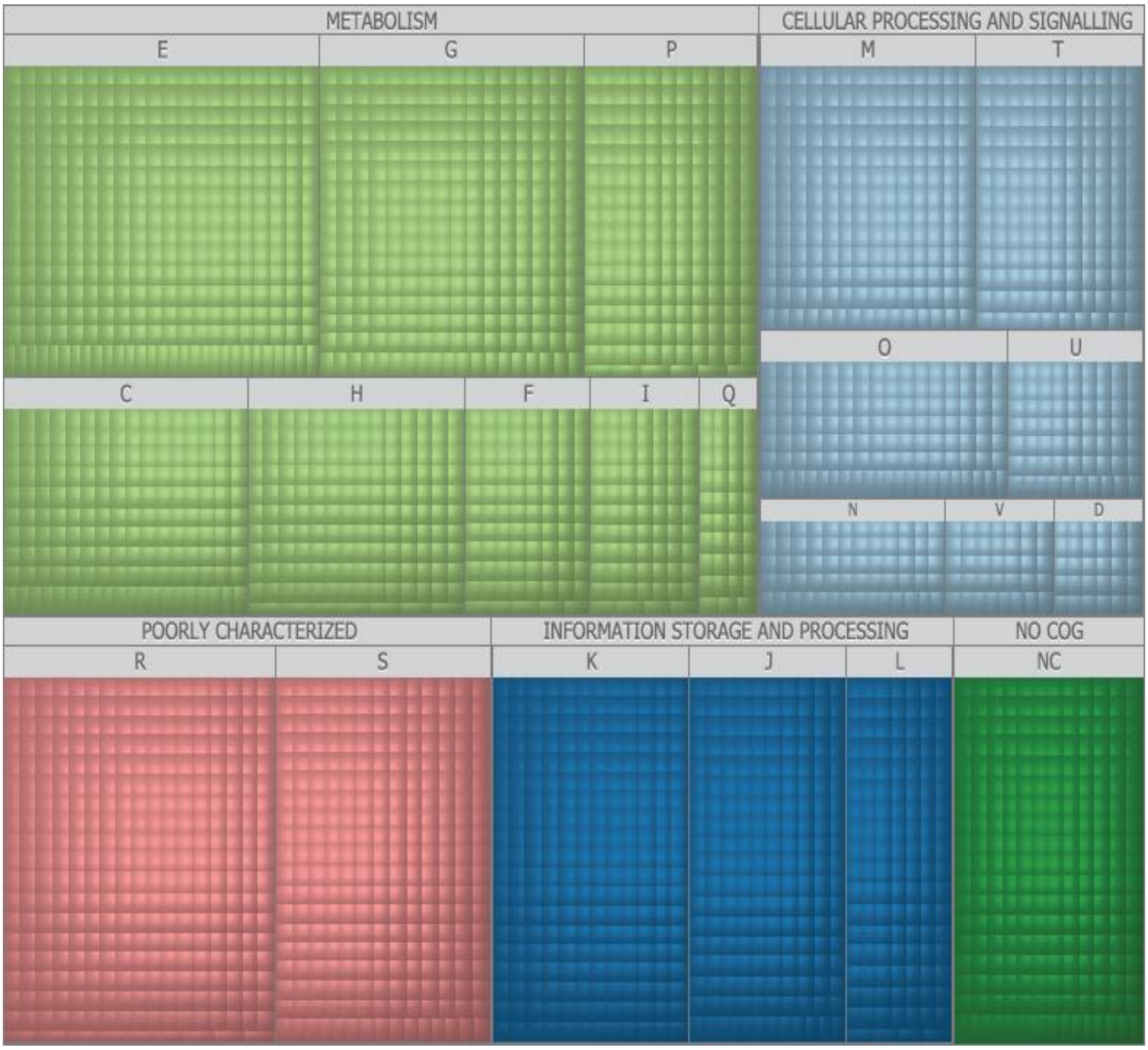


Fig 2 The core genome of *Pantoea*. The core genes are grouped firstly into the general cellular processes (indicated by the different colours) that they play a role in and secondly into the functional categories that they belong to. Some genes have multiple functions and are thus grouped into more than one functional category or cellular process. The abbreviations used correspond with the COG database. Cellular processes and signalling (light blue) is depicted by ‘Cell cycle control, cell division, chromosome partitioning’ (**D**), ‘Cell motility’ (**N**), ‘Cell wall/membrane/envelope biogenesis’ (**M**), Defence mechanisms (**V**),‘Intracellular trafficking, secretion, and vesicular transport’ (**U**), ‘Posttranslational modification, protein turnover, chaperones’ (**O**) and ‘Signal transduction mechanisms’ (**T**). Information storage and processing (dark blue) is depicted by ‘Chromatin structure and dynamics’ (**B**), ‘Replication, recombination and repair’ (**L**), ‘RNA processing and modification’ (**A**), ‘Transcription’ (**K**) and ‘Translation, ribosomal structure and biogenesis’ (**J**). Metabolism (light green) is depicted by ‘Amino acid transport and metabolism’ (**E**), ‘Carbohydrate transport and metabolism’ (**G**), ‘Coenzyme transport and metabolism’ (**H**), ‘Energy production and conversion’ (**C**), ‘Inorganic ion transport and metabolism’ (**P**), ‘Lipid transport and metabolism’ (**I**), ‘Nucleotide transport and metabolism’ (**F**) and ‘Secondary metabolites biosynthesis, transport and catabolism’ (**Q**). Poorly characterized genes (pink) consist of ‘General function prediction only’ (**R**) and ‘Function unknown’ (**S**), with some genes having no related COG descriptions (**N C**) (dark green).

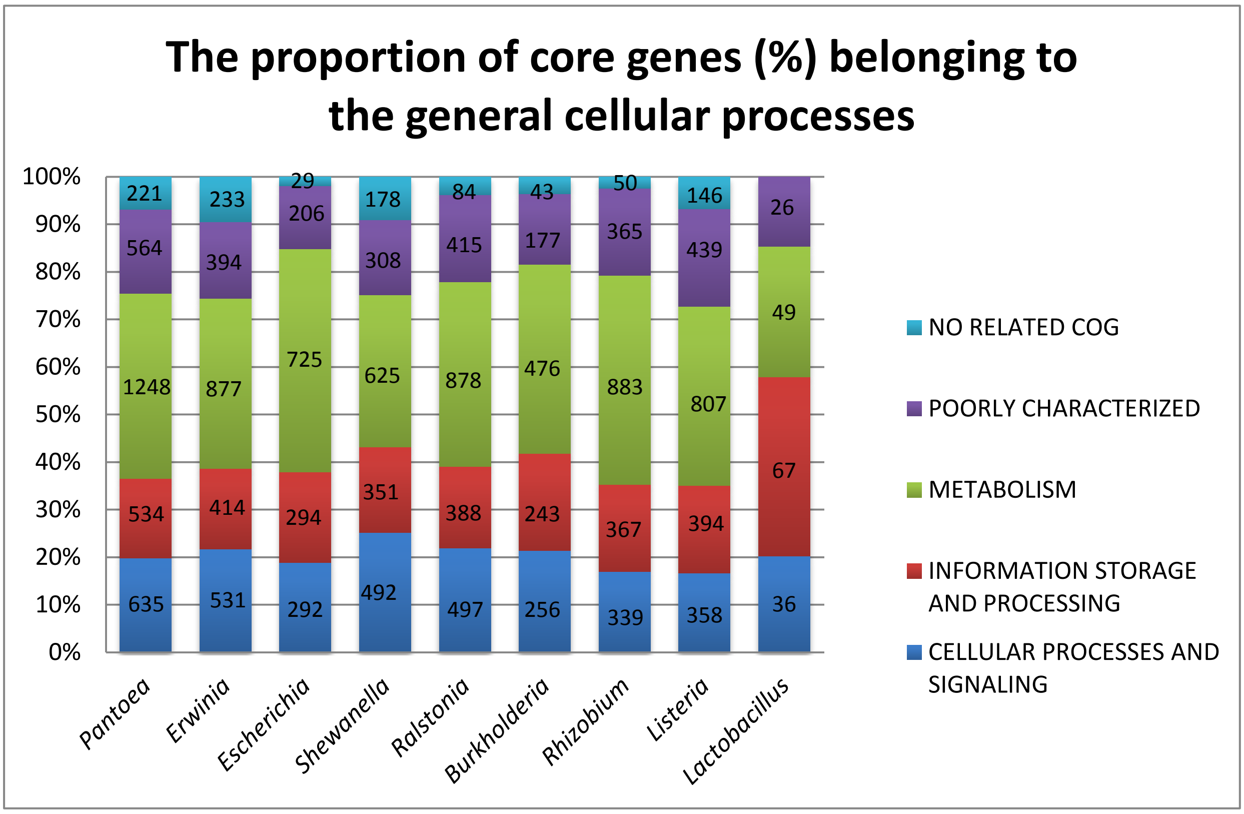


Fig 3 The proportion of core genes of different genera (*Pantoea, Erwinia, Escherichia, Ralstonia, Burkholderia, Shewanella, Rhizobium, Listeria* and *Lactobacillus*) playing a role in each general cellular process. The core genomes of the genera consist of 2907, 2295, 1435, 2074, 1096, 1769, 2389, 1973 and 169 genes respectively. The number of genes representing each cellular process is indicated in each section. Numerous core genes have a dual functionality, and can thus be grouped into more than one general cellular process or functional category.

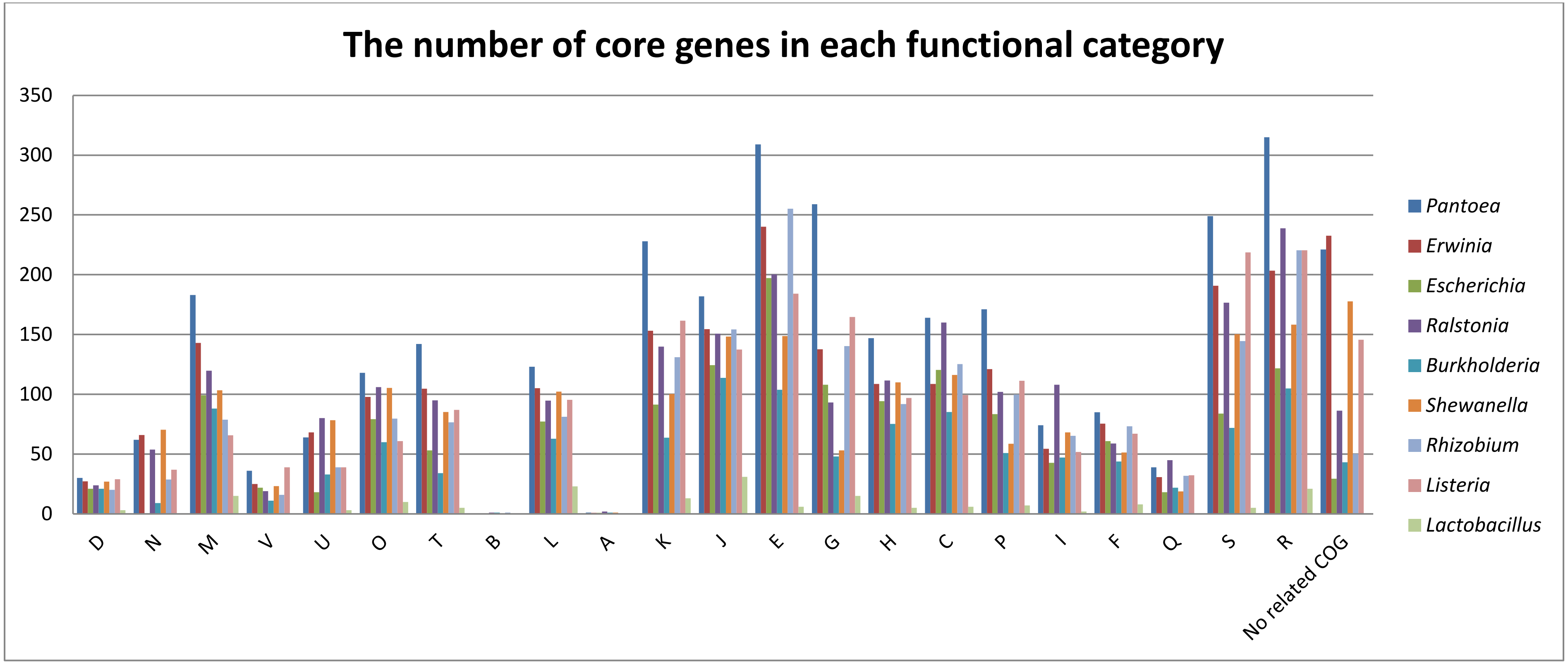


Fig 4 The amount of core genes present in each functional category. The abbreviations used correspond with the COG database. Cellular processes and signalling is depicted by ‘Cell cycle control, cell division, chromosome partitioning’ (**D**), ‘Cell motility’ (**N**), ‘Cell wall/membrane/envelope biogenesis’ (**M**),Defence mechanisms (**V**),‘Intracellular trafficking, secretion, and vesicular transport’ (**U**), ‘Posttranslational modification, protein turnover, chaperones’ (**O**) and ‘Signal transduction mechanisms’ (**T**).Information storage and processing is depicted by ‘Chromatin structure and dynamics’ (**B**), ‘Replication, recombination and repair’ (**L**), ‘RNA processing and modification’ (**A**), ‘Transcription’ (**K**) and ‘Translation, ribosomal structure and biogenesis’ (**J**). Metabolism is depicted by ‘Amino acid transport and metabolism’ (**E**), ‘Carbohydrate transport and metabolism’ (**G**), ‘Coenzyme transport and metabolism’ (**H**), ‘Energy production and conversion’ (**C**), ‘Inorganic ion transport and metabolism’ (**P**), ‘Lipid transport and metabolism’ (**I**), ‘Nucleotide transport and metabolism’ (**F**) and ‘Secondary metabolites biosynthesis, transport and catabolism’ (**Q**). Poorly characterized genes consist of ‘Function unknown’ (**S**), ‘General function prediction only’ (**R**) and genes with no related COG descriptions.

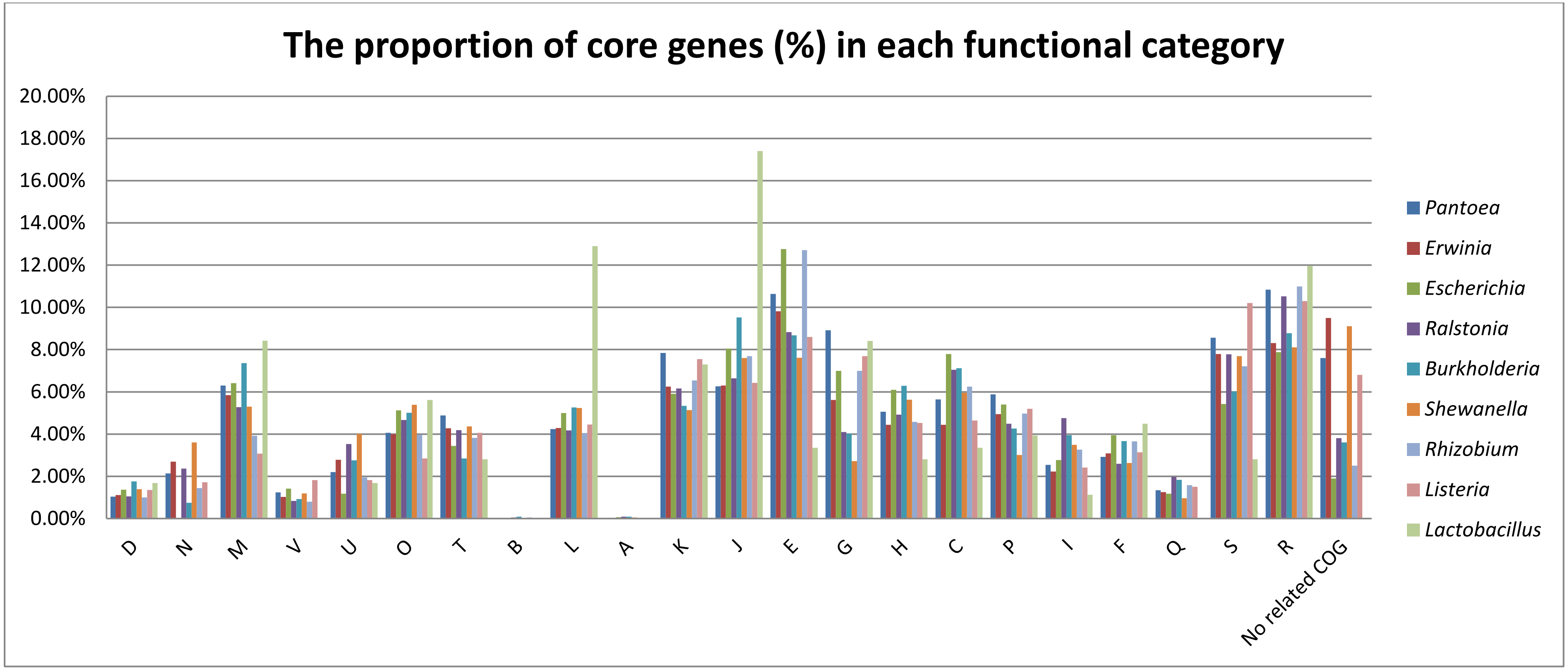


Fig 5 The proportion of core genes from the general cellular processes present in each functional category. The abbreviations used correspond with the COG database. Cellular processes and signalling is depicted by ‘Cell cycle control, cell division, chromosome partitioning’ (**D**), ‘Cell motility’ (**N**), ‘Cell wall/membrane/envelope biogenesis’ (**M**),Defence mechanisms (**V**),‘Intracellular trafficking, secretion, and vesicular transport’ (**U**), ‘Posttranslational modification, protein turnover, chaperones’ (**O**) and ‘Signal transduction mechanisms’ (**T**). Information storage and processing is depicted by ‘Chromatin structure and dynamics’ (**B**), ‘Replication, recombination and repair’ (**L**), ‘RNA processing and modification’ (**A**), ‘Transcription’ (**K**) and ‘Translation, ribosomal structure and biogenesis’ (**J**). Metabolism is depicted by ‘Amino acid transport and metabolism’ (**E**), ‘Carbohydrate transport and metabolism’ (**G**), ‘Coenzyme transport and metabolism’ (**H**), ‘Energy production and conversion’ (**C**), ‘Inorganic ion transport and metabolism’ (**P**), ‘Lipid transport and metabolism’ (**I**), ‘Nucleotide transport and metabolism’ (**F**) and ‘Secondary metabolites biosynthesis, transport and catabolism’ (**Q**). Poorly characterized genes consist of ‘Function unknown’ (**S**) and ‘General function prediction only’ (**R**).

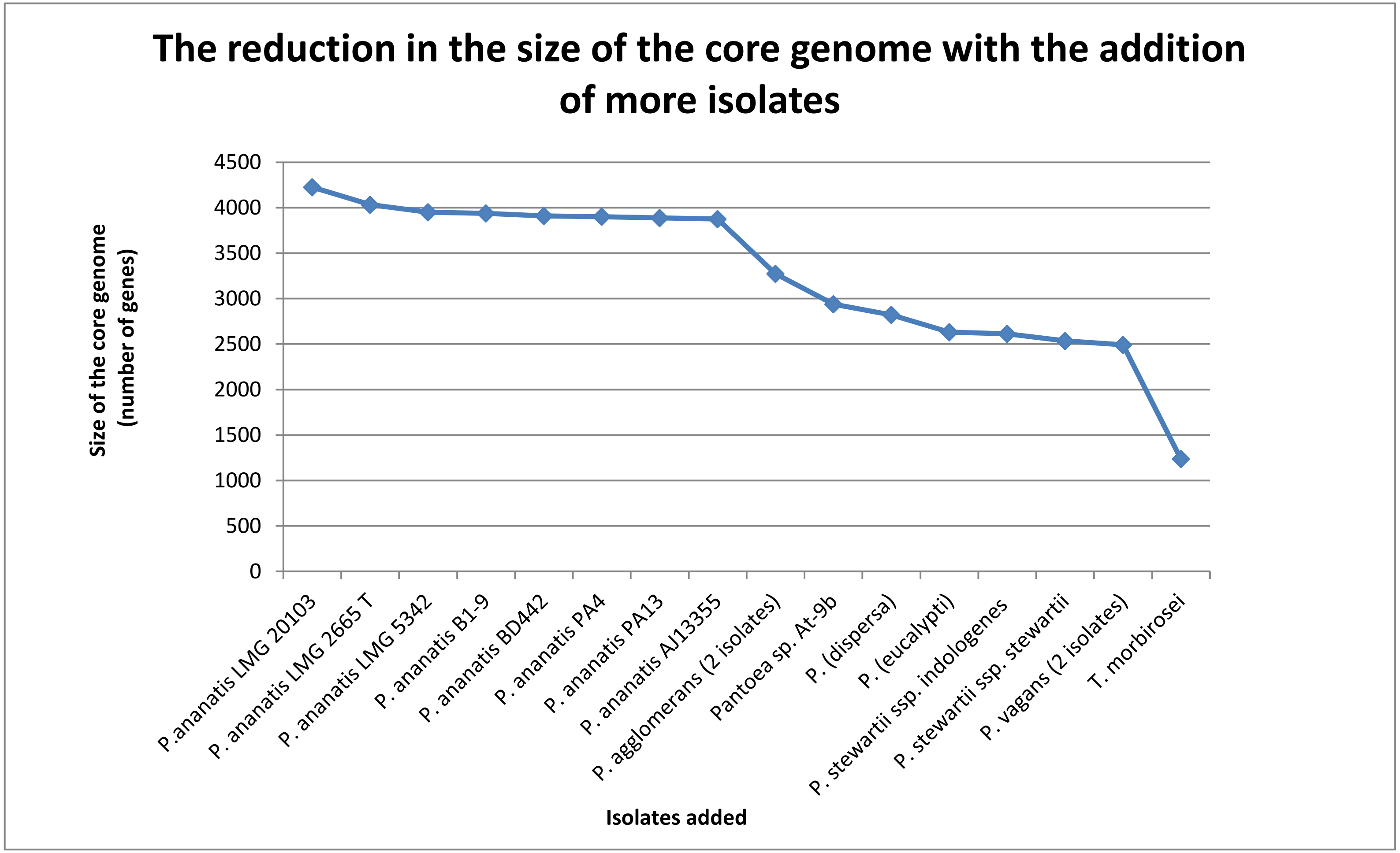


Fig 6 The reduction in the size of the conserved core genome as more divergent isolates are added. The isolates of one species are first added to the comparison, followed by species of the same genus, followed by an isolate of a sister-genus.

Chapter 3

Can the *Pantoea* core genome be used to infer the evolutionary history of the genus?

Can the *Pantoea* core be used to infer the evolutionary history of the genus?

Introduction

Bacterial evolution is driven by both vertical descent and horizontal forces (Woese, 2000) due to the plasticity of the bacterial genome. It is believed, however, that the main signal present in the genome represents that of vertical descent, as the most recent transfer of the majority of genetic information presumably occurs through vertical inheritance (Andam and Gogarten, 2011). This vertical signal can be represented by tree-like structures and have led to the use of species trees for depicting the evolutionary history of organisms (Doolittle and Papke, 2006).

Species trees are most commonly inferred using 16S rRNA gene sequences or by making use of multi-locus sequence analysis (MLSA), which is the phylogenetic analysis of multiple housekeeping genes (Hillis and Dixon, 1991; Gevers et al., 2005; Konstantinidis and Tiedje, 2007). These types of phylogenies are included in most new species descriptions and revisions (Gevers et al., 2005; Rosselló-Móra, 2005). The use of ribosomal multi-locus sequence typing (rMLST), which employ the 53 structural ribosomal protein-coding sequences in an extended MLSA, is also becoming more popular as these genes are found in all groups of bacteria and involves the analysis of far more loci than a typical MLSA (Bennett et al., 2012; Jolley et al., 2012).

Irrespective of the molecular data used, robust phylogenies are not always obtained. One reason for this is the high level of sequence conservation of the commonly used genes, which limits their resolving power among closely related species (Fox et al., 1992; Gevers et al., 2005; Staley, 2006; Richter and Rosselló-Móra, 2009). The marker genes employed may also occur as multiple copies within a single genome (Boucher et al., 2004), thus complicating the identification of orthologous sequences. Furthermore, although horizontal gene transfers (HGT) are thought to occur less frequently with these conserved genes, it has been shown to occur with all genes (Rivera et al., 1998). The use of such horizontally acquired genes for inferring species trees is likely to introduce discordant phylogenetic signals, especially when only a limited number of genes are analysed.

Various authors suggest that the use of a number of genes will increase the reliability of phylogenies (Pamilo and Nei, 1988; Klenk and Göker, 2010). This is because a higher number of characters is thought to improve the ratio between signal and noise (Klenk and Göker, 2010). Similarly, more comprehensive taxon sampling is thought to improve phylogenies by limiting the effects of tree construction artefacts (Hillis, 1998). In other words, reconstruction of a robust species tree is crucially dependent, not only on the number of vertically descended genes utilized, but also by the inclusion of taxa spanning the diversity of the group of organisms under investigation.

The identification of suitable vertically descended markers for inferring species trees may be expedited by the availability of whole-genome sequence (WGS) information for bacterial taxa. This is because bacterial genomes are partitioned into so-called core genomes (i.e. the genetic material present in all members of a taxon) and accessory genomes (i.e. the genetic material that occur in only some members of a taxon) (Lan and Reeves, 2000; Daubin et al., 2002; Tettelin et al., 2008; Mann et al., 2013). Of these, it is the core genome that remains cohesive as it is passed on from parent to progeny (Daubin et al., 2002). This core typically consists of essential genes that are required for survival and are less frequently transferred horizontally than accessory genes (Lan and Reeves, 2000; Hacker and Carniel, 2001; Coenye et al., 2005; Hacker et al., 2012). This hypothesis is supported by many congruent topologies of gene trees constructed from core genes (Daubin et al., 2002), implying that these genes share the same evolutionary path. The most realistic way of constructing a robust species tree, indicating the dominant signal of vertical descent, is thus by using the cohesive core genome (Daubin et al., 2002).

A number of studies have looked at the use of conserved, shared genes (core genomes) for phylogenetic inference of species trees (Daubin et al., 2002; Dutilh et al., 2008; Segata and Huttenhower, 2011), although the available WGS data are biased toward better known species and species with clinical relevance (Dutilh et al., 2008; Segata and Huttenhower, 2011). The phylogenies obtained for core genes by Segata et al., (2011) correlated well to the 16S rRNA phylogenies for these taxa, but only a small number of core genes were used at the different taxonomic ranks as the isolates compared were often very distantly related. Dutihl et al., (2008) used a subset of the conserved core genomes and termed these genes signature genes. These signature genes were those present in all daughter lineages of a clade and also restricted to the clade in question (Dutilh et al., 2008). The number of genes used for each clade was unfortunately limited as these isolates were often also very distantly related and different signature genes had to be used for every clade.

In the current study, the availability of WGS data for a number of species in the genus *Pantoea* were exploited to address genus- and species-based questions using phylogenomic analyses of the conserved core genome. The genus *Pantoea* serves well as a model genus due to the ecological diversity observed within the genus. Members of this genus range from plant pathogens (such as *P. ananatis* (Coutinho and Venter, 2009) and *P. stewartii* ssp. *stewartii* (Frutchey, 1936;

Mergaert et al., 1993) to human pathogens (*P. brenneri* and *P. septica* (Brady et al., 2010a)), with some plant associated species with the ability to infect humans opportunistically (De Champs et al., 2000; De Baere et al., 2004; Cruz et al., 2007; Cataño et al., 2012). There are also isolates that are used for plant-growth promotion (Kim et al., 2012) and bio-control against other bacterial pathogens (Pusey, 2002). Although this wide variety of lifestyles and diversity exist within *Pantoea*, the genus is well defined and appears to be cohesive (Brady et al., 2010b).

The specific aim of this study was to utilize the core genome information for *Pantoea* to infer a robust species tree for the genus. To evaluate the potential impact of HGT on the *Pantoea* core gene set, the overall species tree was also compared to those inferred from different subsets of functionally related core genes. Congruence of the *Pantoea* core gene trees with our current understanding of the systematics and biology of this genus was also investigated by comparison with the commonly employed approaches for inferring species phylogenies (i.e. 16S rRNA analysis, MLSA and rMLST). By using all the shared genes in the genus for phylogenetic analysis, we hope to improve our understanding of the core genome in bacteria, whether it remains cohesive in terms of its evolutionary history and can be used to infer a robust species tree. This will also shed light on the suitability of the currently employed molecular methods used for inferring species trees in bacteria and its correlation to core genome phylogenies.

Materials and experimental methods

Genomes used for analysis

A total of 26 *Pantoea* genomes obtained from various sources, were included in this study (Table 1). These genomes were from closely related isolates and species, as well as distantly related species. The inter- and intraspecies variability present within the genus *Pantoea* was well represented and a good phylogenetic distribution of isolates across the genus was used. A genome sequence for the sister genus *Tatumella* was also included and used for outgroup purposes (Table 1).

Preparation of nucleotide datasets for the commonly employed molecular markers

Datasets consisting of the available 16S rRNA and MLSA nucleotide alignments with the type strains of all species in the genus *Pantoea* was compiled. The MLSA alignments included the partial gene sequences of *atpD*, *gyrB*, *infB* and *rpoB* (as used by Brady et al. [2012]). The gene sequences of the isolates with available genome sequences were identified using BioEdit 7.0.9 (Hall, 1999) with BLAST 2.2.25+ (Altschul et al., 1997) as an accessory application. These sequences were then manually aligned to the existing alignments. The gene sequences of all isolates were truncated to the same length as those utilized by Brady et al. (2012).

Preparation of core genome datasets

The putative core genome of *Pantoea*, as determined previously (Chapter 2 of this dissertation), was used to identify genes shared further by the additional *Pantoea* genomes not used in the previous study (i.e. *P. brenneri* LMG5343^T, *P. cypripedii* LMG2657^T, *P. eucrina* LMG5346^T, *P. septica* LMG5345^T, *Pantoea* sp. A4, *Pantoea* sp. GM01, *Pantoea* sp. Sc-1 and *P. wallisii* LMG26277^T) and that of the outgroup taxon, *Tatumella morbirosei* LMG23360^T. Protein-coding sequences for the additional genomes used were identified using the bacterial gene finding tool, FGENESB (Solovyev and Salamov, 2011), from the Softberry platform available at http://linux1.softberry.com/berry.phtml. Only single copy genes were included in the core genome dataset. A custom python script was used to assign the protein sequences, based on *P. ananatis* AJ13355, to individual FASTA sequence files. This was followed by BLAST searches of the core set of protein sequences to the protein sets of the other taxa and addition of the sequences with the best hits to the sequence files. The desktop version of BLAST 2.2.25+ (Altschul et al., 1997) was used for the identification of homologous genes amongst the different genomes.

The sequence files were then checked in BioEdit 7.0.9 to ensure the presence of gene sequences representing the 27 taxa in each sequence file. For files with truncated sequences or missing taxa, the missing sequences were manually identified by BLAST searches against the protein sequence sets of the full genomes of the relevant taxa. These searches were performed with BLAST 2.2.25+ (Altschul et al., 1997). Complete sequence files were then submitted to the online alignment tool MAFFT v. 7 (Katoh et. al., 2002), after which the alignments were manually inspected and edited (where needed) in BioEdit 7.0.9. Based on the alignments, genes with missing data at the end or beginning of the sequences, possibly due to annotation differences, were truncated to the point where a similar sized dataset was available for all taxa.

Datasets were partitioned into 5 subsets based on Clusters of Orthologous Group (COGs) classification (Tatusov et al., 2000). These included the broad functional groups ‘Cellular processing and signalling’, ‘Information storage and processing’, ‘Metabolism’, ‘Poorly characterized’ and genes with no COG homologues (No COG). The 53 structural ribosomal proteins used for rMLST (Jolley et al., 2012) was also identified as a subset of the ‘Information storage and processing’ set. The sequences in each of these groups were concatenated to produce a data matrix for each of the broad COG groups and the rMLST sequences. Concatenations were performed with

Geneious 6.2 (Drummond et. al., 2011). The complete set of core genes was also concatenated to produce a concatenated core data matrix.

Phylogenetic analyses

Maximum Likelihood (ML) analyses of the nucleotide datasets were performed with MEGA 6 (Tamura et. al., 2013), available at http://www.megasoftware.net/. For both the 16S rRNA gene datasets and the MLSA datasets, these analyses utilized the best-fit substitution models determined with JModelTest (Posada, 2008) using the Akaike Information Criterion (AIC) (Akaike, 1973), which correlated to the previously determined models by Brady et al. (2008). Branch support was estimated using bootstrap analysis based on 1000 replicates.

ML analysis of the core gene data was performed using RAxML 7.2.8 (Stamatakis, 2006), available at http://github.com/stamatak/standard-RAxML. These analyses utilized best-fit amino acid substitution models that were determined with ProtTest 3.2 (Abascal et al., 2005) using AIC (Akaike, 1973) to select the best model. Partitioning of the concatenated data matrices was done manually using the best-fit amino acid model for each partition, together with the appropriate model parameters, to produce a partitioning file for RAxML. Branch support was estimated using 1000 bootstrap replicates and the rapid bootstrapping algorithm followed by the rapid hill-climbing algorithm (Stamatakis et al., 2007). Internode certainty (IC) was also calculated for the concatenated core and COG trees, using the best likelihood tree and the bootstrap trees for each analysis in RAxML 7.4.6, available at http://github.com/stamatak/standard-RAxML (Stamatakis, 2006; Salichos and Rokas, 2013).

Dendroscope 3 (Huson and Scornavacca, 2012), available from http://ab.inf.uni-tuebingen.de/data/software/ dendroscope3/download/welcome.html, was used for visualization and editing of phylogenetic trees. Tests of alternative tree topologies were conducted using the version of the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa, 1999) as implemented in RAxML 7.2.8 (Stamatakis, 2006). These SH-like tests utilized the topologies inferred with a single isolate of each species from the core and core subset (COG) trees as well as from the published MLSA tree (Brady et al., 2012) and the ribosomal subset. Although the nodes of the published MLSA tree and the ribosomal tree were not well supported by bootstrapping, the topologies of the most likely trees were used for this analysis. The constructed 16S rRNA gene tree and the MLSA tree were excluded from the analysis due to very low support at the majority of the nodes.

Results

Core genes and their functions

The genes used for analysis were those present in all 27 genomes of the isolates included in the study (*Pantoea* isolates with genome data available and the *Tatumella morbirosei* isolate). One-thousand, two-hundred and seventy-five (1275) conserved single-copy protein-coding genes could be extracted from all the sequenced genomes. Approximately 30% of the genes were truncated at the beginning or the end of the sequence as annotation differences occurred or because these genes were at the start or the end of contiguous regions. These alignments were truncated to the point where all taxa had data available. It is possible that some of the genes present in all 27 isolates were excluded as some genes could have been missed due to the fact that the genomes used were not fully assembled and consisted of several contiguous segments.

Of the 1275 extracted protein sequences, the majority was involved in ‘Metabolism’ (515), followed by ‘Information processing and signalling’ (289) and then by ‘Cellular processing and signalling’ (264). There were 196 genes of this conserved core that was still ‘Poorly characterized’ and 60 genes that showed no match to sequences in the COG database (No COG homologues) (Chapter 2 of this dissertation).

Phylogenetic trees

The 16S rRNA gene tree had very low (below 70%) or no bootstrap support at the majority of the nodes (Figure 1). The undescribed isolates consistently grouped with the same isolates in the MLSA (Figure 2). There was, however, very little resolution in the trees as the lack of support at the majority of the nodes led to the collapse of these branches.

The MLSA tree (Figure 2) had more nodes with significant bootstrap support than the 16S rRNA gene tree. The two major phylogenetic groups (group 1 and 2; Figure 2) were recovered with high bootstrap support, although there were many deeper nodes without bootstrap support. The first group included *P. agglomerans* to *P. brenneri* and the second group included *P. stewartii* to *P. ananatis*. The constructed MLSA tree consisting of the type strains and the isolates included in this study (Figure 2) differ from the published tree (Brady et al., 2012) in the branching order of the rest of the taxa, including *P. cypripedii* through to *P.dispersa*, as well as the positions of *P. septica* and *P. deleyi*. However, the backbone of the MLSA tree had limited support. Based on the consistent grouping of the undescribed environmental isolates with the specific type strains, *Pantoea* sp. aB was used to represent *P.*

eucalypti, *Pantoea* sp. SL1-M5 was used to represent a *P. agglomerans* isolate, *Pantoea* sp. Sc-1 was used to represent *P. anthophila* and the *Plautia stali* symbiont was used to represent *P. dispersa*. The three other undescribed isolates did not group sufficiently close to any of the type strains to represent a specific species and were considered to belong to undescribed *Pantoea* species.

The rMLST tree inferred from the 53 structural ribosomal protein sequences had very low bootstrap support at most nodes (Figure 3). The only exceptions were those representing the two clusters of *P. ananatis* (Group 2; Figure 2) and *P. agglomerans* (Group 1; Figure 2) and their related species. In addition, to this lack of resolution, the rMLST tree also had a different topology when compared to the core and all the COG topologies (see below), as *P. septica* grouped separately from the rest of the species.

Generally, all nodes of the trees obtained in the concatenated analyses of all core genes and the COG groups showed high bootstrap support with the majority of the nodes having bootstrap values of above 90% (Figures 4 - 9). The only exceptions were the node where *P. eucalypti* (*Pantoea* sp. aB) grouped with *P. agglomerans* (Figures 5 & 6) instead of with *P. vagans* (Figure 4, 7, 8 & 9), and in the group containing *P. wallisii*, *P. eucrina, Pantoea* sp. GM01*, P. cypripedii, Pantoea* sp. At-9b and the *Plautia stali* symbiont (*P. dispersa*) (Group 3; Figure 4) (Figures 8 & 9).

Where tree topologies were the same in terms of branching order, slight differences were observed in the branch lengths between trees. Nevertheless, the phylogeny obtained from all the core sequences (core topology) (Figure 4) reflected the same topology in terms of branching order as that of the ‘Metabolism’ phylogeny (Figure 7). The phylogenies obtained for ‘Cellular processing and signalling’ (cellular) (Figure

5) and for ‘Information storage and processing’ (information) (Figure 6) shared the same topology in terms of branching order. The difference between the trees based on these major COG groups in terms of branching order, is the position of *P. eucalypti* (*Pantoea* sp. aB), where the core topology and the metabolism topology shows *P. eucalypti* (*Pantoea* sp. aB) as a sister taxon to *P. vagans* (Figures 4 and 7), and the cellular topology and the information topology shows *P. eucalypti* as a sister taxon of *P. agglomerans* (Figures 5 and 6). The bootstrap support for this branching pattern in the cellular phylogeny and the information phylogeny was however low, with the bootstrap value at this node in the information phylogeny being below 70%. The phylogeny for the sequences with ‘No COG homologues’ (No COG) (Figure 8) and the ‘Poorly characterized’ (poorly characterized) phylogeny (Figure 9) both had different topologies to the previously mentioned topologies with regard to the third group.

Internode certainty

Internode Certainty values were calculated with the bootstrap trees (Salichos and Rokas, 2013) for the concatenated core and COG trees. IC was calculated using the number of trees supporting the most prevalent bipartition and the second most prevalent conflicting bipartition at a specific node. High bootstrap values at nodes with relatively low IC values, indicated that a conflicting split at the node was present in the bootstrap dataset, which occurred frequently in the remaining bootstrap trees and did not support the split observed in the best tree. This means that the next best supported split following the one observed for the “best” tree, was actually well represented in the remaining bootstrap trees. This was observed in the ‘Cellular processing and signalling’ and ‘Poorly characterized’ trees (Figures 5 and 9) at the node where *P. eucalypti* (*Pantoea* sp. aB) grouped with *P. agglomerans* and *P.*

vagans, respectively. All other nodes with high bootstrap support also had high IC values. Values were excluded at nodes where bootstrap values were below 70%, as these IC values were very low, suggesting that the two most prevalent splits were almost equally well supported from the set of bootstrap trees. An example of this scenario was in the 'Information storage and processing' tree at the node where *P. eucalypti* (*Pantoea* sp. aB) grouped with *P. agglomerans*. The bootstrap value for this node was below 70% with an IC value of 0.06, meaning that the alternate splits for this node observed in the collection of bootstrap trees were almost equally well represented.

SH-like tests

The SH-like tests in RAxML 7.2.8 allowed for the evaluation of a set of alternative topologies against each data matrix. These topologies were the topologies obtained from the ML analyses of the core and different COG data matrices and the data matrices used were the core set of genes and the individual COG subsets of the core. These tests provided an indication of whether the tested topologies were scored significantly worse in their likelihoods for the given data matrix than the tree obtained from the data matrix (Figure 10). The hypothesis tested was that alternate topologies were significantly worse than the obtained topologies for each data matrix. The tests were performed at a significance level of 1% (p-value of 0.01, confidence interval of 99%). However, alternative topologies could receive the highest likelihood scores and perform better in the SH-like tests compared to the tree that was created based on that specific dataset. This is possible as these likelihood analyses are not as exhaustive as in the individual analyses and the parameters are not optimised for every topology.

For the core data matrix, the tree topologies for metabolism and the poorly characterized data matrices were not significantly worse, with all other topologies being significantly worse for the core data matrix. The poorly characterized topology as well as the core topology was not significantly worse than the obtained topology for the metabolism data matrix with all other topologies scoring significantly worse. For the cellular dataset, the topologies for the core, metabolism, information and No COG topologies did not score significantly worse than the obtained topology. All topologies except the inferred MLSA topology were also not significantly worse for the information dataset. The cellular, information and inferred MLSA topologies were not significantly worse for the No COG data matrix with all other topologies scoring significantly worse than the obtained tree. All alternate topologies were significantly worse for the dataset for the poorly characterized genes. The core, metabolism, cellular, information and poorly characterized topologies were not significantly worse than the obtained topology for the ribosomal dataset. The inferred MLSA topology (Brady et al., 2012) was thus seen as significantly worse for all datasets except the No COG data matrix. The core topology was thus not significantly worse than the obtained topologies for most datasets.

Discussion

The use of core genome data provides a complete sample for genes thought to be acquired through vertical descent (Hacker and Carniel, 2001; Daubin et al., 2002). Phylogenetic analysis of this set of genes allowed reconstruction a robust species tree for the genus. This tree had high bootstrap and IC support values at all nodes and was not significantly worse than the best trees inferred from most of the COG data matrices. Therefore, the species tree inferred from the core genome information for *Pantoea* represents a robust phylogenetic hypothesis against which to evaluate what is known regarding the evolutionary history of the species in this genus.

Comparison of the various phylogenetic trees constructed in this study suggested some correlation between the level of bootstrap support for the trees and the size of the data matrix used. The larger datasets with more molecular data (core and COG), produced trees with better supported nodes, which is consistent with expectations (Pamilo and Nei, 1988; Hillis et al., 1994; Klenk and Göker, 2010). This was even more evident when the core genome and COG trees were compared to the commonly employed molecular markers. The 16S rRNA gene tree, MLSA tree and the rMLST tree all had insufficient resolution, especially at deeper nodes. Due to the low support for nodes in these trees, they cannot be reliably used to investigate the evolution of the genus. Although the limitations of 16S rRNA and MLSA markers for inferring species trees has been noted before (Konstantinidis and Tiedje, 2005b; Gevers et al., 2005; Richter and Rosselló-Móra, 2009), rMLST has been suggested to be invaluable for reliably tracing the phylogenetic relationships among members of a bacterial genus (Bennett et al., 2012; Jolley et al., 2012). In the current study, the topology of the *Pantoea* rMLST were markedly different from those inferred from any of the core gene datasets, which was also evident from the results of the SH-like tests. This suggests that, like 16S rRNA analysis and MLSA, rMLST is also not suitable avenue for inferring a species tree for *Pantoea*. The reliability of rMLST would thus need to be tested on each taxonomic group before employing it as a method for obtaining a species tree.

Topology differences were observed for the different COG trees. It is generally expected that the core genome remains fully cohesive (Hacker and Carniel, 2001; Daubin et al., 2002; Coenye et al., 2005), thus the same evolutionary history was expected for all core genes. Although differences were observed between the COG trees, almost all nodes of these trees were very well supported. The high support for nodes may suggest cohesion between the genes in a specific COG group, meaning that the majority of the genes in those COG groups share the same evolutionary history, although the evolutionary history may differ between the different COG groups.

The differences among the phylogenies inferred from the larger COG datasets (information, cellular and metabolism) were limited to a single node. The low bootstrap support at that node (*P. agglomerans*, *P. vagans* and *P. eucalypti*) in the cellular and information phylogenies, together with the number of alternate topologies that were not significantly worse for these data matrices, indicate that the difference in topology between these datasets might be due to a lack of resolution in the more conserved datasets (information and cellular) rather than different evolutionary histories (Galtier and Daubin, 2008). However, the possibility that this incongruence is due to incomplete lineage sorting (Salichos and Rokas, 2013; Galtier and Daubin, 2008) cannot be excluded as these species have likely diverged relatively recently (unpublished data). This indicates that the conserved metabolic genes may in fact be necessary to obtain a reliable, well supported phylogeny, especially to differentiate between closely related species, as the informational genes may not give the required resolution between these species.

Large conflicts were observed between the COG trees for the smaller data matrices (No COG homologues and the poorly characterized datasets) and the COG trees for the larger data matrices. The alternate topologies were mostly significantly worse for the smaller data matrices, indicating that these phylogenies were notably different from the other phylogenies obtained. The genes with no homologues in the COG database (No COG) may be unique to the genus *Pantoea*, and although these genes are conserved in the genus, it may not share its evolutionary history with the majority of core genes (Rivera et al., 1998). This may be caused by HGT of these genes, selection or through other evolutionary processes (Doyle, 1992; Galtier and Daubin, 2008). The genes that are poorly characterized may also have been acquired laterally or selected for to evolve in a specific manner (Doyle, 1992; Galtier and Daubin, 2008), as these genes are conserved in a number of other, often distantly related, species, giving a different phylogenetic picture to that obtained for the larger COG groups.

The use of core genomes for phylogenomic analyses have proven to be a useful tool for obtaining species trees of bacteria (Daubin et al., 2002; Galtier and Daubin, 2008; Dutilh et al., 2008; Segata and Huttenhower, 2011) and provides better supported and robust phylogenies compared to the commonly employed molecular markers. It has, however, been suggested that the use of only a few genes with strong phylogenetic signal may be more feasible, but their identification will be difficult without the use of a robust phylogeny for comparison (Konstantinidis et al., 2006a; Salichos and Rokas, 2013). The results presented here indicates that the choice of core genes for analysis remains problematic as different subsets of the core provide different evolutionary pictures, as has been suggested before (Rivera et al., 1998; Jain et al., 1999). In *Pantoea*, however, the larger COG groups of genes like information, cellular and metabolism genes do, however, appear to remain cohesive, while HGT appears to have impacted upon only some of the core genome subsets. The overall *Pantoea* core genome tree together with those inferred from the various subsets of core genes will thus be invaluable for addressing questions pertaining to the evolutionary history of bacterial genus (Heath et al., 2008), with regard to not only the core genome but also the accessory genome. This may, in future, shed light on how different traits, like pathogenicity have evolved in different lineages and how these traits were distributed across species.

References

ABASCAL, F., ZARDOYA, R. & POSADA, D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics,* 21**,** 2104-2105.

ADAMS, A. S., JORDAN, M. S., ADAMS, S. M., SUEN, G., GOODWIN, L. A., DAVENPORT, K. W., CURRIE, C. R. & RAFFA, K. F. (2011) Cellulose-degrading bacteria associated with the invasive woodwasp Sirex noctilio. *ISME Journal,* 5**,** 1323-1331.

AKAIKE, H. (1973) Information theory and an extension of the maximum likelihood principle. *Second international symposium on information theory.* Akademinai Kiado.

ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25**,** 3389-3402.

ANDAM, C. P. & GOGARTEN, J. P. (2011) Biased gene transfer in microbial evolution. *Nature Reviews Microbiology,* 9**,** 543-555.

BENNETT, J. S., JOLLEY, K. A., EARLE, S. G., CORTON, C., BENTLEY, S. D., PARKHILL, J. & MAIDEN, M. C. J. (2012) A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology,* 158**,** 1570-1580.

BOUCHER, Y., DOUADY, C. J., SHARMA, A. K., KAMEKURA, M. & DOOLITTLE, F. W. (2004) Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *Journal of Bacteriology,* 186**,** 3980-3990.

BRADY, C. L., CLEENWERCK, I., VAN DER WESTHUIZEN, L., VENTER, S. N., COUTINHO, T. A. & DE VOS, P. (2012) *Pantoea rodasii* sp. nov., *Pantoea rwandensis* sp. nov. and *Pantoea wallisii* sp. nov., isolated from Eucalyptus. *International Journal of Systematic and Evolutionary Microbiology,* 62**,** 1457-1464.

BRADY, C., CLEENWERCK, I., VENTER, S., VANCANNEYT, M., SWINGS, J. & COUTINHO, T. (2008) Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Systematic and Applied Microbiology,* 31, 447-460.

BRADY, C. L., CLEENWERCK, I., VENTER, S. N., ENGELBEEN, K., DE VOS, P. & COUTINHO, T. A. (2010a) Emended description of the genus *Pantoea*, description of four species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov., and transfer of *Pectobacterium cypripedii* (Hori 1911) Brenner et al. 1973 emend. Hauben et al. 1998 to the genus as *Pantoea cypripedii* comb. nov. *International Journal of Systematic and Evolutionary Microbiology,* 60**,** 2430-2440.

BRADY, C. L., VENTER, S. N., CLEENWERCK, I., VANDEMEULEBROECKE, K., DE VOS, P. & COUTINHO, T. A. (2010b) Transfer of *Pantoea citrea*, *Pantoea punctata* and *Pantoea terrea* to the genus *Tatumella* emend. as *Tatumella citrea* comb. nov., *Tatumella punctata* comb. nov. and *Tatumella terrea* comb. nov. and description of *Tatumella morbirosei* sp. nov. *International Journal of Systematic and Evolutionary Microbiology,* 60**,** 484-494.

BROWN, S. D., UTTURKAR, S. M., KLINGEMAN, D. M., JOHNSON, C. M., MARTIN, S. L., LAND, M. L., LU, T.-Y. S., SCHADT, C. W., DOKTYCZ, M. J. & PELLETIER, D. A. (2012) Twenty-one genome sequences from *Pseudomonas* species and 19 genome sequences from diverse bacteria isolated from the rhizosphere and endosphere of *Populus deltoides*. *Journal of Bacteriology,* 194**,** 5991-5993.

CATAÑO, J. C., ECHEVERRI, L. M. & SZELA, C. (2012) Bacterial contamination of clothes and environmental items in a third-level hospital in Colombia. *Interdisciplinary Perspectives on Infectious Diseases,* 2012.

CHOI, O., LIM, J. Y., SEO, Y.-S., HWANG, I. & KIM, J. (2012) Complete genome sequence of the rice pathogen *Pantoea ananatis* strain PA13. *Journal of Bacteriology,* 194**,** 531.

COENYE, T., GEVERS, D., VAN DE PEER, Y., VANDAMME, P. & SWINGS, J. (2005) Towards a prokaryotic genomic taxonomy. *Federation of European Microbiological Societies Microbiology Reviews,* 29**,** 147-167.

COUTINHO, T. A. & VENTER, S. N. (2009) *Pantoea ananatis*: an unconventional plant pathogen. *Molecular Plant Pathology,* 10**,** 325-335.

CRUZ, A. T., CAZACU, A. C. & ALLEN, C. H. (2007) *Pantoea agglomerans* - a plant pathogen causing human disease. *Journal of Clinical Microbiology***,** 45 (6), 1989-1992.

DAUBIN, V., GOUY, M. & PERRIÈRE, G. (2002) A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Research,* 12**,** 1080-1090.

DE BAERE, T., VERHELST, R., LABIT, C., VERSCHRAEGEN, G., WAUTERS, G., CLAEYS, G. & VANEECHOUTTE, M. (2004) Bacteremic infection with *Pantoea ananatis*. *Journal of Clinical Microbiology,* 42**,** 4393-4395.

DE CHAMPS, C., LE SEAUX, S., DUBOST, J. J., BOISGARD, S., SAUVEZIE, B. & SIROT, J. (2000) Isolation of *Pantoea agglomerans* in two cases of septic monoarthritis after plant thorn and wood sliver injuries. *Journal of Clinical Microbiology,* 38**,** 460-461.

DE MAAYER, P., CHAN, W. Y., REZZONICO, F., BÜHLMANN, A., VENTER, S. N., BLOM, J., GOESMANN, A., FREY, J. E., SMITS, T. H. M., DUFFY, B. & COUTINHO, T. A. (2012) Complete genome sequence of clinical isolate *Pantoea ananatis* LMG 5342. *Journal of Bacteriology,* 194**,** 1615-1616.

DE MAAYER, P., CHAN, W. Y., VENTER, S. N., TOTH, I. K., BIRCH, P. R. J., JOUBERT, F. & COUTINHO, T. A. (2010) Genome sequence of *Pantoea ananatis* LMG20103, the causative agent of Eucalyptus blight and dieback. *Journal of Bacteriology,* 192**,** 2936-2937.

DOOLITTLE, F. W. & PAPKE, T. R. (2006) Genomics and the bacterial species problem. *Genome Biology,* 7**,** 116.

DOYLE, J. J. (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany,* 17**,** 144-163.

DRUMMOND, A. J., ASHTON, B., BUXTON, S., CHEUNG, M., COOPER, A., DURAN, C., FIELD, M., HELED, J., KEARSE, M. & MARKOWITZ, S. (2011) Geneious v5. 4. Available at http://www.geneious.com

DUTILH, B. E., SNEL, B., ETTEMA, T. J. G. & HUYNEN, M. A. (2008) Signature genes as a phylogenomic tool. *Molecular Biology and Evolution,* 25**,** 1659-1667.

FOX, G. E., WISOTZKEY, J. D. & JURTSHUK, P. J. (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *International Journal of Systematic Bacteriology,* 42**,** 166-170.

FRUTCHEY, C. W. (1936) A study of stewarts disease of sweet corn caused by *Phytomonas stewartii,* Michigan State College, Agricultural Experiment Station, Michigan, USA.

GALTIER, N. & DAUBIN, V. (2008) Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 363**,** 4023-4029.

GEVERS, D., COHAN, F. M., LAWRENCE, J. G., SPRATT, B. G., COENYE, T., FEIL, E. J., STACKEBRANDT, E., VAN DE PEER, Y., VANDAMME, P., THOMPSON, F. L. & SWINGS, J. (2005) Re-evaluating prokaryotic species. *Nature Reviews,* 3**,** 733-739.

HACKER, J. R. & CARNIEL, E. (2001) Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO reports,* 2**,** 376-381.

HACKER, J. R. H., DOBRINDT, U. & KURTH, R. (2012) *Genome plasticity and infectious diseases*, ASM Press, Washington, USA.

HALL, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series.*

HARA, Y., KADOTANI, N., IZUI, H., KATASHKINA, J., KUVAEVA, T., ANDREEVA, I., GOLUBEVA, L., MALKO, D., MAKEEV, V., MASHKO, S. & KOZLOV, Y. (2012) The complete genome sequence of *Pantoea ananatis* AJ13355, an organism with great biotechnological potential. *Applied Microbiology and Biotechnology,* 93**,** 331-341.

HEATH, T. A., HEDTKE, S. M. & HILLIS, D. M. (2008) Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution,* 46**,** 239-257.

HILLIS, D. M. (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology,* 47**,** 3-8.

HILLIS, D. M. & DIXON, M. T. (1991) Ribosomal DNA: Molecular Evolution and Phylogenetic Inference. *The Quarterly Review of Biology,* 66**,** 411-453.

HILLIS, D. M., HUELSENBECK, J. P. & CUNNINGHAM, C. W. (1994) Application and accuracy of molecular phylogenies. *Science,* 264**,** 671-676.

HONG, K.-W., GAN, H. M., LOW, S.-M., LEE, P. K. Y., CHONG, Y.-M., YIN, W.-F. & CHAN, K.-G. (2012) Draft genome sequence of *Pantoea* sp. strain A4, a *Rafflesia*-associated bacterium that produces N-acylhomoserine lactones as quorum-sensing molecules. *Journal of Bacteriology,* 194**,** 6610.

HUSON, D. H. & SCORNAVACCA, C. (2012) Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology,* 61**,** 1061-1067.

JAIN, R., RIVERA, M. C. & LAKE, J. A. (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences,* 96**,** 3801-3806.

JOLLEY, K. A., BLISS, C. M., BENNETT, J. S., BRATCHER, H. B., BREHONY, C., COLLES, F. M., WIMALARATHNA, H., HARRISON, O. B., SHEPPARD, S. K., CODY, A. J. & MAIDEN, M. C. J. (2012) Ribosomal Multi-Locus Sequence Typing: universal characterisation of bacteria from domain to strain. *Department of Zoology.* Oxford, University of Oxford.

KATOH, K., MISAWA, K., KUMA, K. I. & MIYATA, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research,* 30**,** 3059-3066.

KIM, H. J., LEE, J. H., KANG, B. R., RONG, X., MCSPADDEN GARDENER, B. B., JI, H. J., PARK, C.-S. & KIM, Y. C. (2012) Draft genome sequence of *Pantoea ananatis* B1-9, a nonpathogenic plant growth-promoting bacterium. *Journal of Bacteriology,* 194**,** 729.

KLENK, H. P. & GÖKER, M. (2010) En route to a genome-based classification of Archaea and Bacteria? *Systematic and Applied Microbiology,* 33**,** 175-182.

KOBAYASHI, H., KAWASAKI, K., TAKEISHI, K. & NODA, H. (2011) Symbiont of the stink bug *Plautia stali* synthesizes rough-type lipopolysaccharide. *Microbiological Research,* 167**,** 48-54.

KONSTANTINIDIS, K. T., RAMETTE, A. & TIEDJE, J. M. (2006) The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 361**,** 1929-1940.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. (2005b) Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology,* 187**,** 6258-6264.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. (2007) Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Current Opinion in Microbiology,* 10**,** 504-509.

LAN, R. & REEVES, P. R. (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends in Microbiology,* 8**,** 396-401.

LANAVE, C., PREPARATA, G., SACONE, C. & SERIO, G. (1984) A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution,* 20**,** 86-93.

MANN, R. A., SMITS, T. H. M., BÜHLMANN, A., BLOM, J., GOESMANN, A., FREY, J. E., PLUMMER, K. M., BEER, S. V., LUCK, J., DUFFY, B. & RODONI, B. (2013) Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. *PLoS ONE,* 8**,** e55644.

MEDRANO, E. G. & BELL, A. A. (2012) Genome sequence of *Pantoea* sp. Strain Sc 1, an opportunistic cotton pathogen. *Journal of Bacteriology,* 194**,** 3019.

MERGAERT, J., VERDONCK, L. & KERSTERS, K. (1993) Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., respectively, and description of *Pantoea stewartii* subsp. *indologenes* subsp. nov. *International Journal of Systematic Bacteriology,* 43**,** 162-173.

PAMILO, P. & NEI, M. (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution,* 5**,** 568-583.

POSADA, D. (2008) jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution,* 25**,** 1253-1256.

PUSEY, P. L. (2002) Biological control agents for fire blight of apple compared under conditions limiting natural dispersal. *Plant Disease,* 86**,** 639-644.

RICHTER, M. & ROSSELLÓ-MÓRA, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences,* 106**,** 19126-19131.

RIVERA, M. C., JAIN, R., MOORE, J. E. & LAKE, J. A. (1998) Genomic evidence for two functionally distinct gene classes. *Proceedings of the National Academy of Sciences,* 95**,** 6239-6244.

ROSSELLÓ-MÓRA, R. (2005) Updating prokaryotic taxonomy. *Journal of Bacteriology,* 187**,** 6255-6257.

SALICHOS, L. & ROKAS, A. (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature,* 497**,** 327-333.

SEGATA, N. & HUTTENHOWER, C. (2011) Toward an efficient method of identifying core genes for evolutionary and functional microbial phylogenies. *PLoS ONE,* 6 (9)**,** e24704.

SHIMODAIRA, H. & HASEGAWA, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution,* 16**,** 1114-1116.

SOLOVYEV, V. & SALAMOV, A. (2011) Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics and its applications in agriculture, biomedicine and environmental studies***,** 61-78.

STALEY, J. T. (2006) The bacterial species dilemma and the genomic-phylogenetic species concept. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 361**,** 1899-1909.

STAMATAKIS, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics,* 22**,** 2688-2690.

STAMATAKIS, A., BLAGOJEVIC, F., NIKOLOPOULOS, D. S. & ANTONOPOULOS, C. D. (2007) Exploring new search algorithms and hardware for phylogenetics: RAxML meets the IBM Cell. *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology,* 48**,** 271-286.

SUEN, G., SCOTT, J. J., AYLWARD, F. O., ADAMS, S. M., TRINGE, S. G., PINTO-TOMÁS, A. N. A., FOSTER, C. E., PAULY, M., WEIMER, P. J., BARRY, K. W., GOODWIN, L. A., BOUFFARD, P., LI, L., OSTERBERGER, J., HARKINS, T. T., SLATER, S. C., DONOHUE, T. J. & CURRIE, C. R. (2010) An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genetics,* 6**,** e1001129.

TAMURA, K. & NEI, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution,* 10**,** 512-526.

TATUSOV, R. L., GALPERIN, M. Y., NATALE, D. A. & KOONIN, E. V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research,* 28**,** 33-36.

TETTELIN, H., RILEY, D., CATTUTO, C. & MEDINI, D. (2008) Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology,* 11**,** 472-477.

WOESE, C. R. (2000) Interpreting the universal phylogenetic tree. *Proceedings of the National Academy of Sciences,* 97**,** 8392-8396.

Table 1 Information of isolates used in this study

Isolate	**Geographic origin**	**Isolated from (host)**	**Lifestyle**	**Reference (Genome sequence)**
Pantoea agglomerans E325	USA	'Gala' apple blossoms	Biocontrol agent	*[1]
Pantoea agglomerans MP2	South Africa	Termite	Symbiont	*[2]
Pantoea agglomerans SL1-M5	USA	Woodwasp	Symbiont	(Adams et al., 2011)
Pantoea ananatis AJ13355	Japan	Soil	Saprophyte	(Hara et al., 2012)
Pantoea ananatis B1-9	Korea	Onion	Plant-growth promoter	(Kim et al., 2012)
Pantoea ananatis BD442	South Africa	Maize	Pathogen	#
Pantoea ananatis LMG20103	South Africa	Eucalyptus	Pathogen	#(De Maayer et al., 2010)
Pantoea ananatis LMG2665 T	Brazil	Pineapple	Pathogen	#
Pantoea ananatis LMG5342	USA	Human	Pathogen	# (De Maayer et al., 2012)
Pantoea ananatis PA4	South Africa	Onion seed	Pathogen	#
Pantoea ananatis PA13	Korea	Rice	Pathogen	(Choi et al., 2012)
Pantoea anthophila Sc-1	USA	Cotton	Pathogen	(Medrano and Bell, 2012)
Pantoea brenneri LMG5343 T	USA	Human (Urethra)	Pathogen	#
Pantoea cypripedii LMG2657 T	USA	Orchid	Pathogen	#
Pantoea sp. At-9b	USA	Leaf cutter ant	Free-living	(Suen et al., 2010)
Pantoea dispersa (*Plautia stali* symbiont)	Japan	Stink bug(*Plautia stali*)	Symbiont	(Kobayashi et al., 2011)
Pantoea eucalypti aB	USA	Bark beetle	Free-living	NCBI (Unpublished): ASM17965v1
Pantoea eucrina LMG5346 T	USA	Human (Trachea)	Pathogen	#
Pantoea septica LMG5345 T	USA	Human (Stool)	Pathogen	#
Pantoea sp. A4	Malaysia	Rafflesia flower	Endophyte	(Hong et al., 2012)
Pantoea sp. GM01	USA	Poplar	Endophyte	(Brown et al., 2012)
Pantoea stewartii subsp. *indologenes* LMG2632 T	India	Millet	Pathogen	#
Pantoea stewartii subsp. *stewartii* DC283	USA	Maize	Pathogen	*[3]
Pantoea vagans C9-1	USA	'Jonathan' apple (fruit)	Biocontrol agent	*[1]
Pantoea vagans MP7	South Africa	Termite	Symbiont	*[2]
Pantoea wallisii LMG26277 T	South Africa	Eucalyptus	Pathogen	#
Tatumella morbirosei LMG23360 T	Philippines	Pineapple	Pathogen	#

\# Genomes sequenced by the Department of Microbiology and Plant Pathology, University of Pretoria.

* Genome sequences obtained from collaborators ([1]Plant Protection Division, Agroscope Changins-Wädenswil, Switzerland; [2]Department of Biology, University of Copenhagen, Denmark; [3]Genome Evolution Laboratory, University of Wisconsin-Madison, USA).

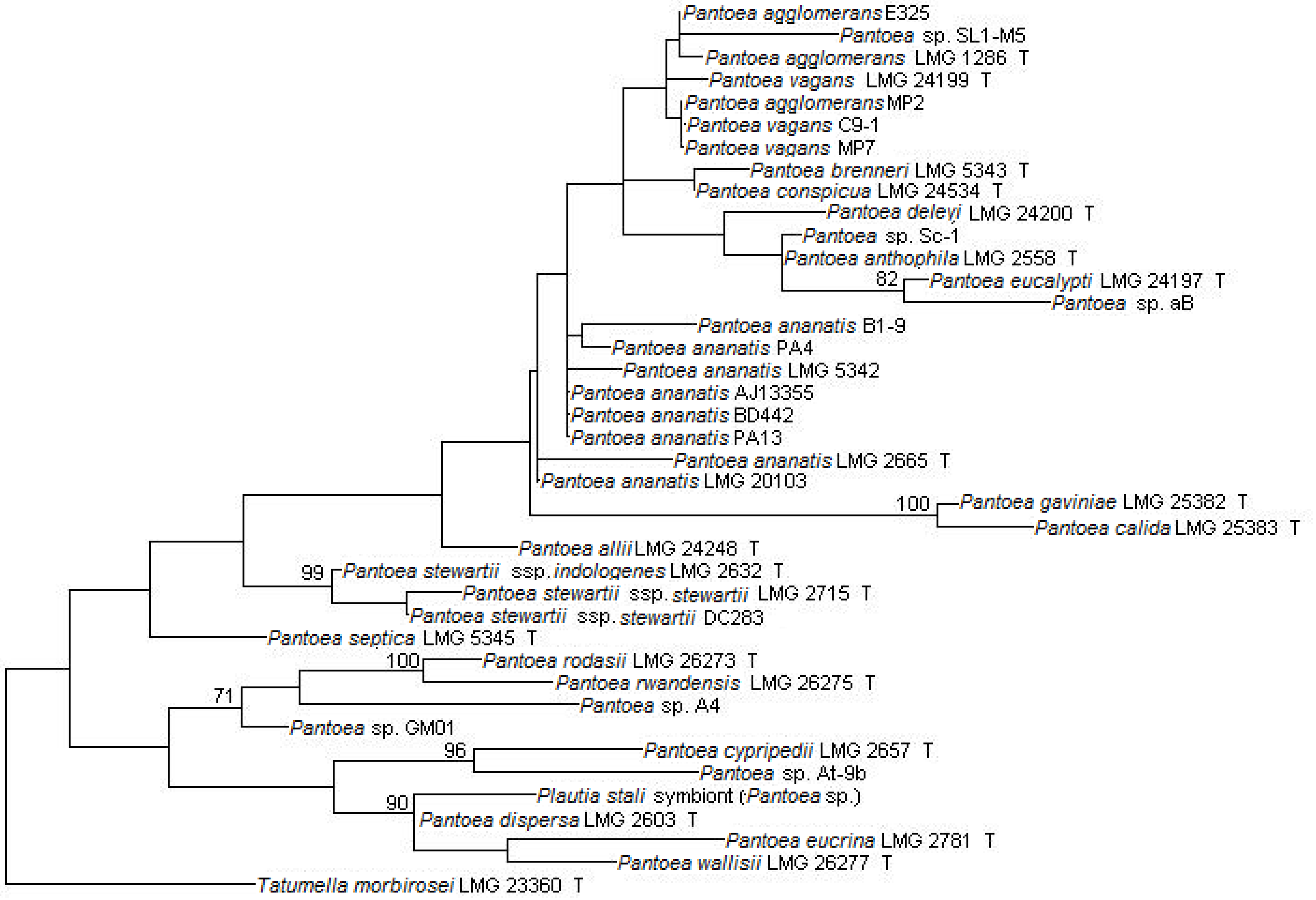


Figure 1 Maximum likelihood (ML) tree constructed from the 16S rRNA gene sequence of all described type strains in the genus *Pantoea* as well as the additional isolates used in this study (not type strain isolates). The Tamura-Nei nucleotide substitution model (Tamura and Nei, 1993) was used for the likelihood analysis with estimation of the proportion of invariable sites and the gamma distribution. Bootstrapping was also performed with 1000 bootstrap replicates. The bootstrap values are indicated as node labels. Bootstrap values below 70% were excluded.

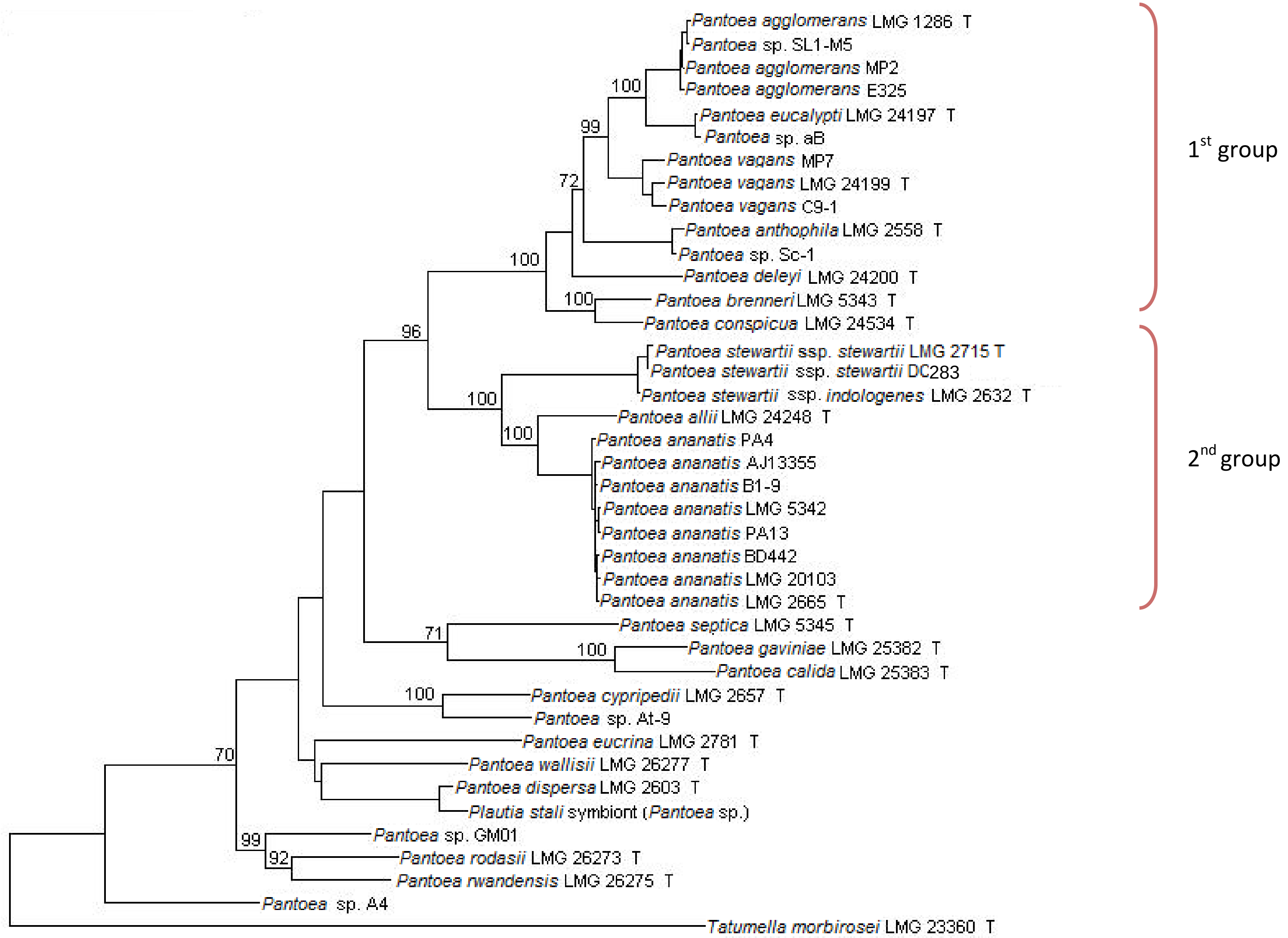


Figure 2 ML tree from the genes currently used in multi-locus sequence analysis (MLSA) for the description of species in the genus *Pantoea*. The genes used were *atpD*, *gyrB*, *infB* and *rpoB* (Brady et al., 2008). Type strains as well as the additional strains used in the study (not type strain isolates) were included in the analysis. The general time reversible (GTR) nucleotide substitution model (Lanave et al., 1984) was used with an estimation of the proportion of invariable sites and the gamma distribution. Bootstrap analysis was also performed with 1000 replicates. Bootstrap values are indicated at nodes. All bootstrap values below 70% were excluded.

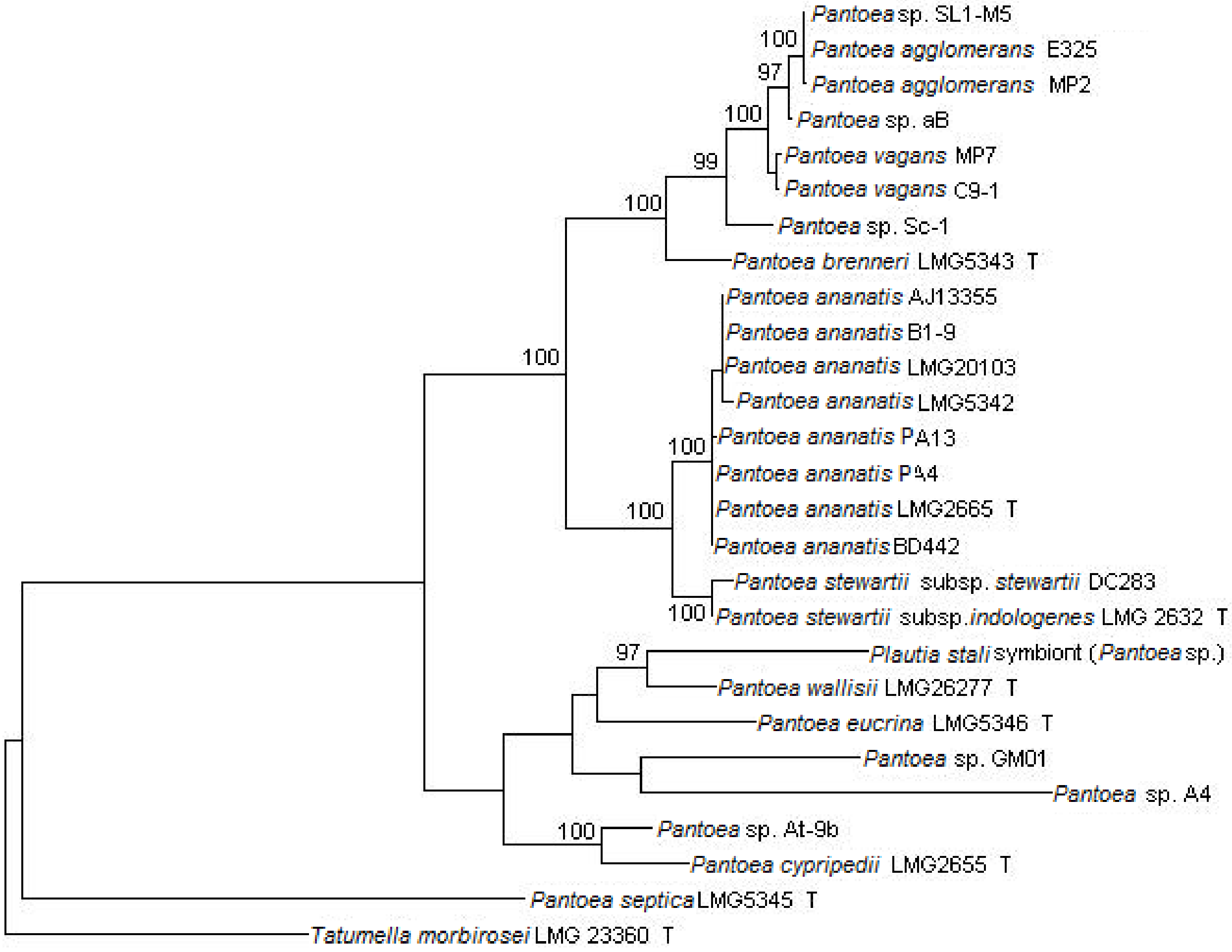


Figure 3 Concatenated ML tree constructed from the 53 structural ribosomal protein sequences used for ribosomal typing (Jolley et al., 2013). The dataset consisted of 53 partitions with the appropriate amino acid substitution models and model parameters for each partition. All bootstrap values below 70% as well as values for within species branches were excluded.

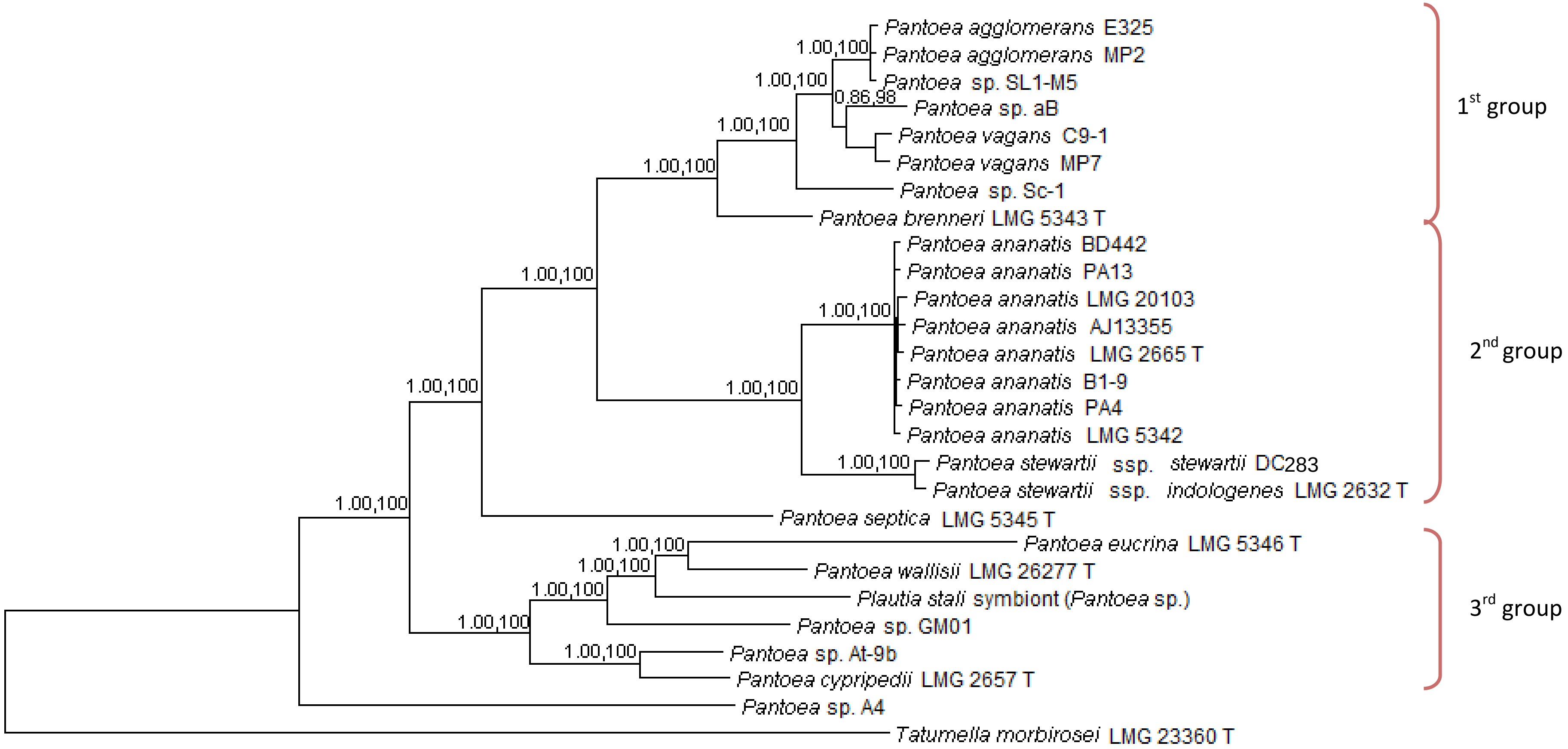


Figure 4 Concatenated ML tree constructed from the protein sequences of all core genes shared between the genus *Pantoea* and the outgroup, *Tatumella morbirosei*. The dataset consisted of a concatenation of 1275 protein sequences. The dataset was partitioned into 1275 partitions, with each partition being assigned the best-fit amino acid substitution model and model parameters. An internode certainty (IC) analysis with the bootstrap trees and a thousand bootstrap replicates were performed (Salichos & Rokas, 2013). Node labels contain both IC and bootstrap values (IC,bootstrap). All bootstrap values were above 70%, thus only bootstrap values for within species branches were excluded.

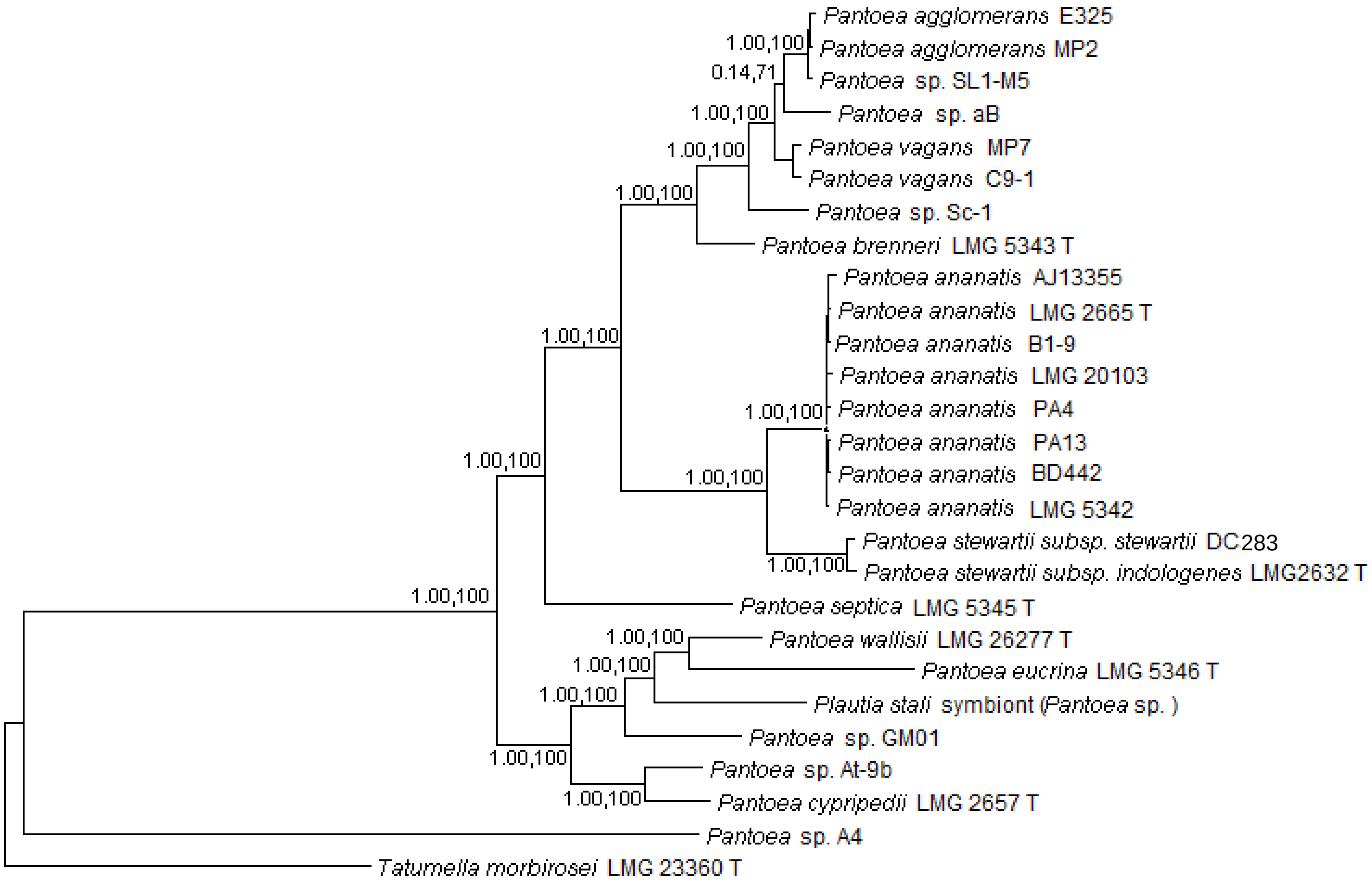


Figure 5 Concatenated ML tree constructed from the protein sequences of all core genes shared by the genus *Pantoea* and the outgroup, *T. morbirosei*, involved in cellular processing and signalling. The dataset consisted of a concatenation of 264 protein sequences. The matrix was partitioned into 264 partitions with the appropriate amino acid substitution model and model parameters being assigned to each partition. Internode certainty (IC) and bootstrap values for 1000 bootstrap replicates are indicated at nodes (IC, bootstrap). All bootstrap values were above 70% and only bootstrap values for within species branches were excluded.

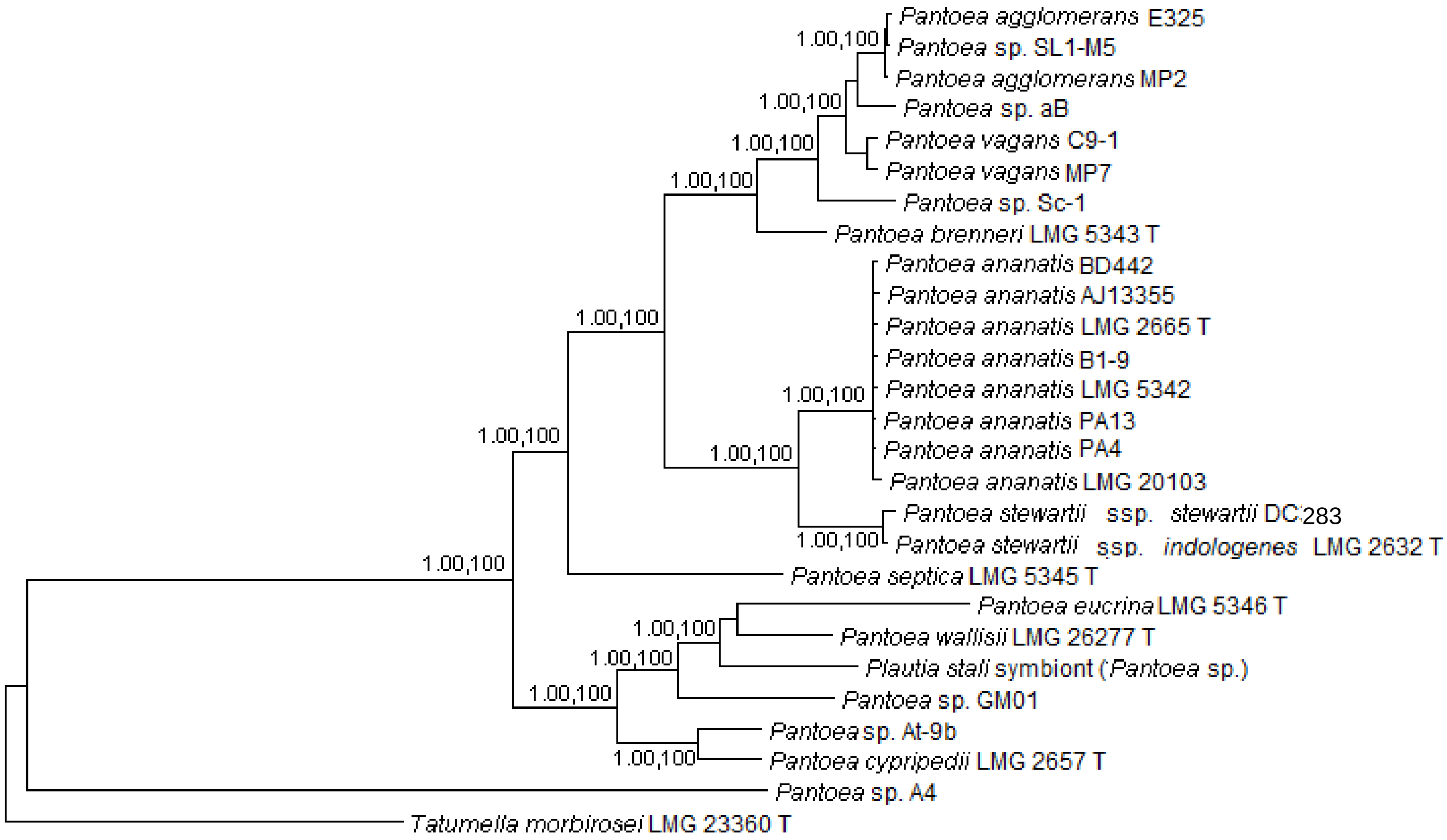


Figure 6 Concatenated ML tree constructed from protein sequences of core genes shared by *Pantoea* and the outgroup, *T. morbirosei*, involved in information storage and processing. The dataset consisted of a concatenation of 289 protein sequences. The dataset was partitioned into 289 partitions with the appropriate amino acid substitution model and model parameters being applied to each partition. Node labels consists of internode certainty (IC) values and bootstrap values for 1000 bootstrap replicates (IC,bootstrap). Bootstrap values below 70% and values for within species branches were excluded.

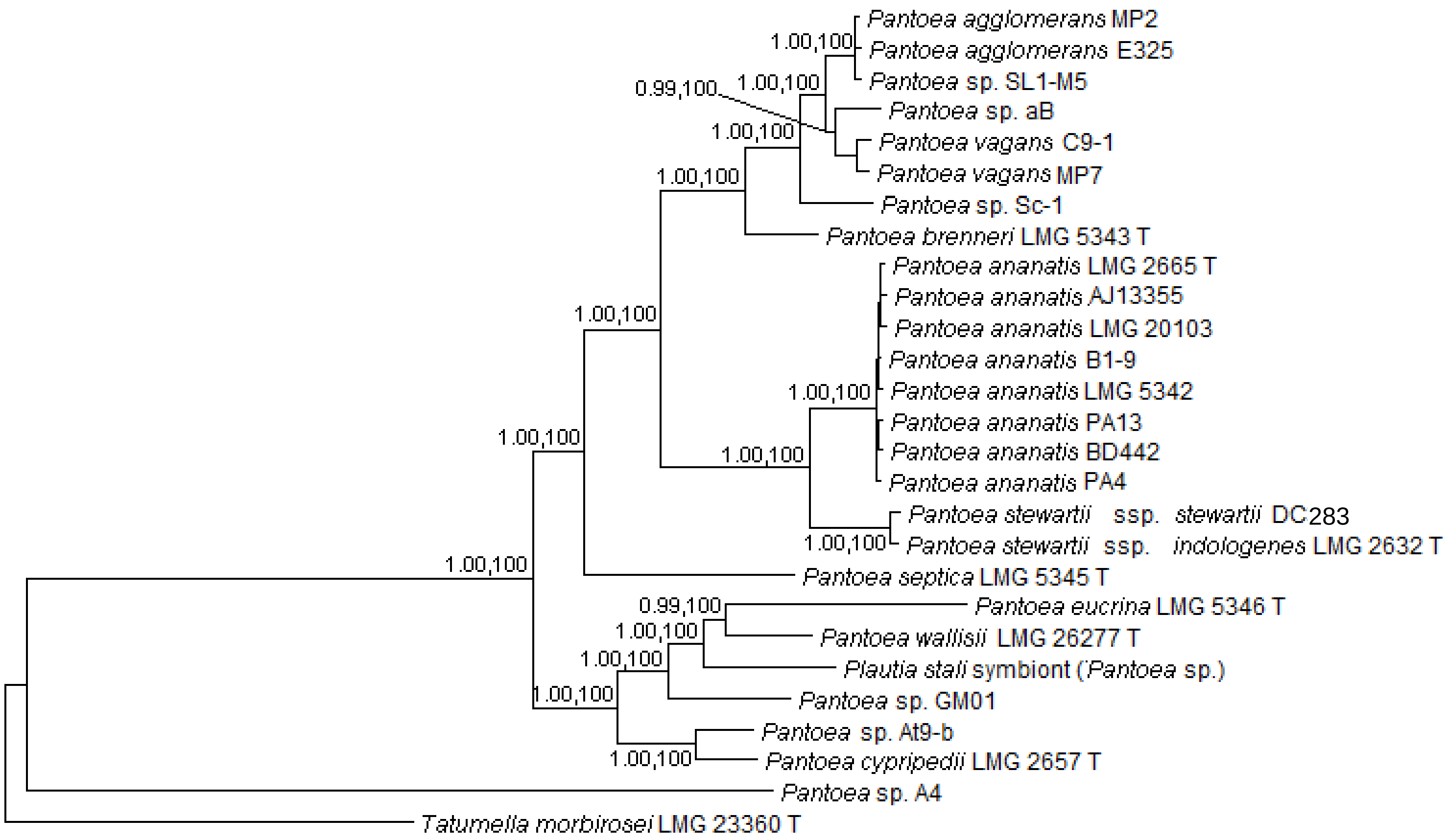


Figure 7 Concatenated ML tree constructed from the protein sequences of all core genes shared by the genus *Pantoea* and the outgroup, *T. morbirosei*, involved in metabolism. The dataset consisted of a concatenation of 515 protein sequences. The dataset was partitioned into 515 partitions with the appropriate amino acid substitution model and model parameters specified for each partition. Internode certainty (IC) and bootstrap values for 1000 bootstrap replicates are indicated at nodes (IC, bootstrap). All bootstrap values were above 70% and bootstrap values for within species branches were excluded.

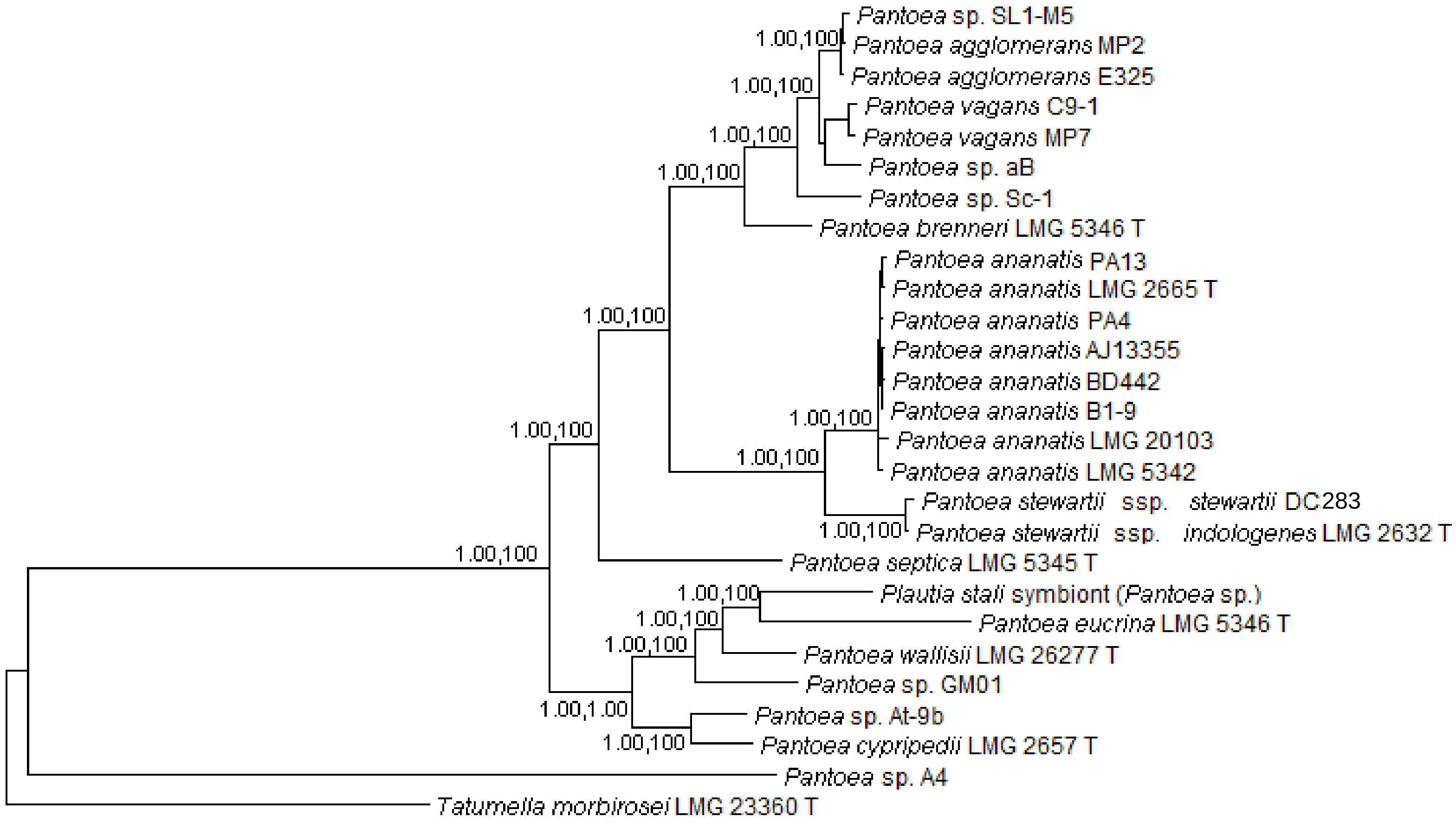


Figure 8 Concatenated ML tree constructed from 60 protein sequences of genes with no COG homologues that are shared by the genus *Pantoea* and the outgroup *T. morbirosei*. The dataset consisted of a concatenation of 60 protein sequences and the data matrix was partitioned with the appropriate amino acid substitution model and model parameters for each partition. Internode certainty (IC) and bootstrap values are indicated at nodes (IC, bootstrap). Bootstrap values below 70% and values for within species branches were excluded.

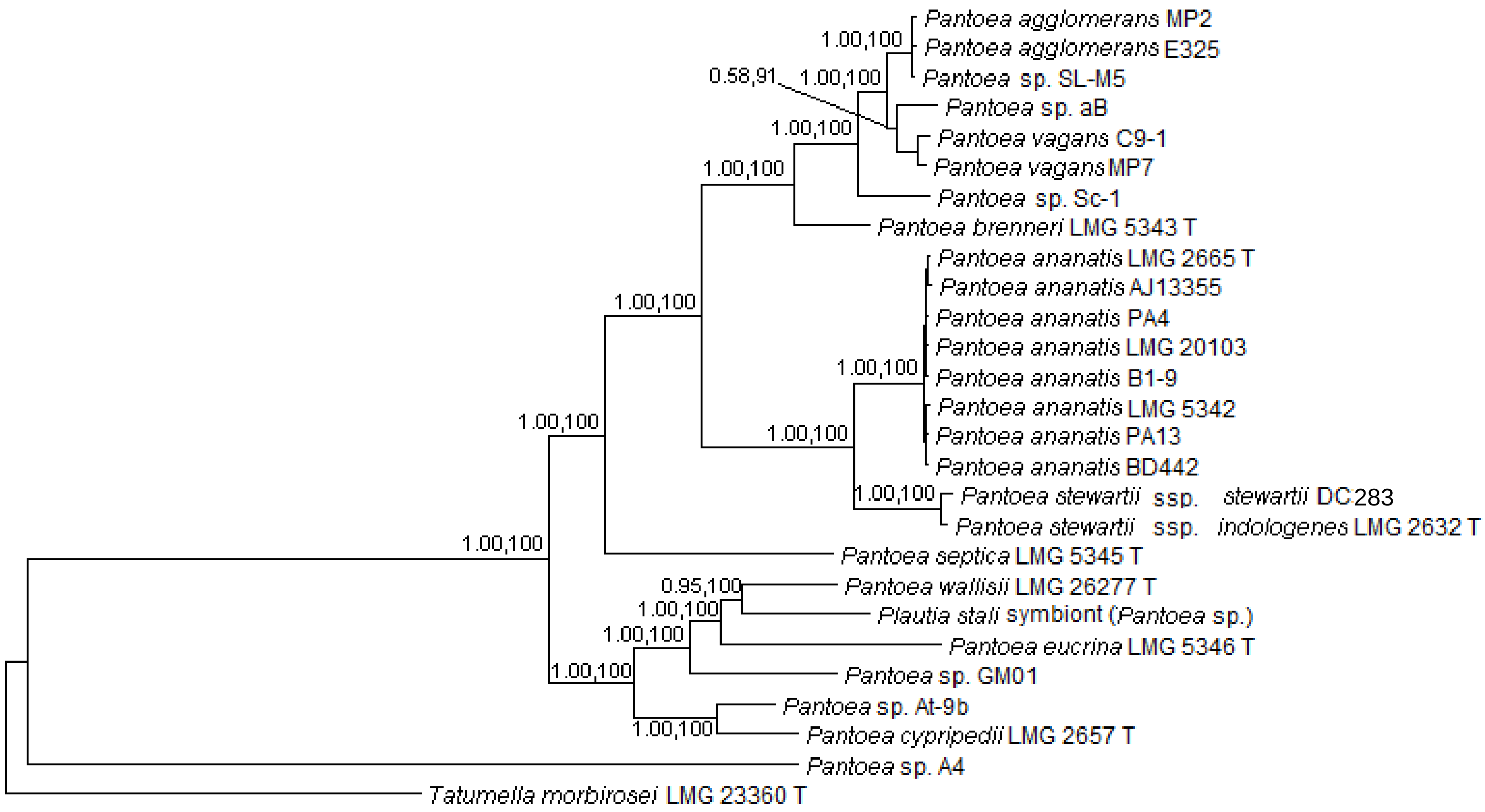


Figure 9 Concatenated ML tree constructed from the protein sequences of all core genes shared by *Pantoea* and the outgroup, *T. morbirosei*, that is still poorly characterized with unknown function or general function predictions. The dataset consisted of a concatenation of 196 protein sequences and was partitioned into 196 partitions with the appropriate amino acid substitution models and model parameters. Internode certainty (IC) and bootstrap values from 1000 bootstrap replicates are indicated at the nodes (IC, bootstrap). Bootstrap values were above 70%, thus only values for within species branches were excluded.

		Datasets						
		Core	Metabolism	Cellular	Information	No COG	Poorly characterized	Ribosomal
Topologies	# of genes in data set	1275	515	264	289	60	196	53
	Core	Obtained/Best	No	Best	Best	Yes	Yes	No
	Metabolism	No	Obtained	No	No	Yes	Yes	No
	Cellular	Yes	Yes	Obtained	No	No	Yes	No
	Information	Yes	Yes	No	Obtained	No	Yes	No
	No COG	Yes	Yes	No	No	Obtained/Best	Yes	Yes
	Poorly characterized	No	Best	Yes	No	Yes	Obtained/Best	No
	rMLST **	Yes	Yes	Yes	No	Yes	Yes	Obtained/Best
	Inferred MLSA **	Yes	Yes	Yes	Yes	No	Yes	Yes

* All results at significance level of 1%

** The topology for the ribosomal and inferred MLSA was used, although low bootstrap support for many deeper branches was observed

Obtained = Tree topology seen as best phylogenetic hypothesis obtained with parameter optimisation and exhaustive ML search;
Best = Tree topology determined as best phylogenetic hypothesis based on the SH-test;
No = Tree topology is not significantly worse than the obtained tree at a p-value of 0.01;
Yes = Tree topology is significantly worse than the obtained tree at a p-value of 0.01;

Figure 10 Matrix of the results of the tree topology comparisons based on Shimodaira-Hasegawa (SH)-like tests as implemented in RAxML 7.2.8. Topologies with only a single isolate for species were compared to eliminate differences in topologies within species as the focus was between species groups. The obtained tree was the best tree obtained for the corresponding dataset in an exhaustive maximum-likelihood analysis with parameter optimisation. Green blocks indicate topologies that are either not significantly worse or better than the obtained tree at a p-value of 0.01. Different tree topologies could be shown to be the best phylogenetic hypothesis as the likelihood search was not as exhaustive as the initial analyses and parameters were not optimised for each tree topology during the SH-like test. Red blocks indicate topologies for the specific datasets that are significantly worse at a p-value of 0.01.

Summary

The delineation of bacterial taxonomic ranks has been shown to be problematic as concepts for the delineation of these groups are lacking. The use of operational criteria was implemented to delineate bacterial species and genera. However, this approach has often caused artificial groupings that do not reflect the same level of diversity present amongst bacterial groups and are not predictive of the phenotype. In this thesis the shared conserved core genome of a genus was identified to determine whether this core can be used to 1) investigate the cohesiveness of a bacterial genus and 2) obtain a robust species tree that that will reflect our current understanding of the evolution of members of this genus.

The genus *Pantoea* was used as a model for this thesis as it is an economically and clinically important genus. The diversity in this genus ranges from plant pathogens to human pathogens, with some plant associated isolates infecting humans opportunistically. There are also isolates that are used as bio-control agents against other plant pathogenic bacteria as well as plant-growth promoting isolates. There are a number of genomes available that give a good representation of the phenotypic and genotypic diversity observed within this genus.

The *Pantoea* core genome was characterised in terms of function using Clusters of Orthologous Groups (COGs) and compared to the core genomes of other bacterial genera. The sequences of the obtained COG groups were thereafter used to construct phylogenomic trees based on different subsets of the core genome as well as on the entire core genome.

In general, the fraction of core genes in each functional category did not differ markedly between most genera. The size of the core genome does however impact on the general ratios of genes in each functional category. Smaller core displayed a larger fraction of genes that are associated with ‘Information storage and processing’, as these genes are essential for survival. This then suggests that the number of genes involved in each general cellular process reduces gradually as the size of the core genome decreases. The genera with a high number of conserved genes, including those usually not considered as essential for survival, maintain these genes as this may be related to the survival of these species in different and varying habitats.

The results obtained suggest a high level of genetic and functional cohesiveness in the core genome of *Pantoea*. This is evident from the considerably larger size of the core genome compared to other genera, the function of the genes conserved and the reduction in core genome size with the addition of sister-taxa. Overall, this supports the hypothesis that the core genome of a bacterial genus could be used to delineate the boundaries of the genus. This may also be able to delineate higher taxa and with the increased in WGS data representative of all bacterial diversity, this approach will become highly feasible.

The full set of shared genes between *Pantoea* and the outgroup taxon was used to obtain a robust phylogeny with very high support at all nodes. With the comparisons of the COG subsets of the shared genes, it was observed that most of the gene groups supported the phylogeny obtained based on the core. The slight differences amongst these trees did not indicate the alternative trees to be significantly worse, based on the Shimodaira-Hasegawa-like tests. There was however significant differences between the groups of the poorly characterised genes and the genes unique to *Pantoea* (No COG homologues). This suggests that although the genes in these COG groups share the same evolutionary history (reflected by the high bootstrap and internode certainty values on the nodes), the groups do not necessarily share the same evolutionary history as the majority of core genes. It was also seen that the informational genes do not necessarily provide the best phylogeny as there is a lack in resolution between closely related species. The metabolism genes seem to play an important role when elucidating the evolutionary relationships between these species.

Appendix A

Table 1 The core genome of the genus *Pantoea* with the Clusters of Orthologous Genes (COG) associations for each core gene

Pantoea gene	General cellular process	Functional category	Letter association	COG class
PANA_0001	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_0002	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_0003	METABOLISM	Carbohydrate transport and metabolism	G	COG0524
PANA_0004	METABOLISM	Carbohydrate transport and metabolism	G	COG1879
PANA_0005	METABOLISM	Carbohydrate transport and metabolism	G	COG1172
PANA_0006	METABOLISM	Carbohydrate transport and metabolism	G	COG1129
PANA_0008	METABOLISM	Carbohydrate transport and metabolism	G	COG1869
PANA_0009	POORLY CHARACTERIZED	Function unknown	S	COG3111
PANA_0010	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_0010	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_0011	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_0012	METABOLISM	Inorganic ion transport and metabolism	P	COG3158
PANA_0013	POORLY CHARACTERIZED	General function prediction only	R	COG0714
PANA_0014	POORLY CHARACTERIZED	General function prediction only	R	COG2425
PANA_0015	POORLY CHARACTERIZED	Function unknown	S	COG1262
PANA_0016	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1522
PANA_0017	METABOLISM	Energy production and conversion	C	COG0716
PANA_0018	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0445
PANA_0019	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0357
PANA_0020	METABOLISM	Energy production and conversion	C	COG3312
PANA_0021	METABOLISM	Energy production and conversion	C	COG0356
PANA_0022	METABOLISM	Energy production and conversion	C	COG0636
PANA_0023	METABOLISM	Energy production and conversion	C	COG0711
PANA_0024	METABOLISM	Energy production and conversion	C	COG0712
PANA_0025	METABOLISM	Energy production and conversion	C	COG0056
PANA_0026	METABOLISM	Energy production and conversion	C	COG0224
PANA_0027	METABOLISM	Energy production and conversion	C	COG0055
PANA_0028	METABOLISM	Energy production and conversion	C	COG0355
PANA_0029	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1207
PANA_0030	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0449
PANA_0032	METABOLISM	Inorganic ion transport and metabolism	P	COG0226
PANA_0033	METABOLISM	Inorganic ion transport and metabolism	P	COG0573
PANA_0034	METABOLISM	Inorganic ion transport and metabolism	P	COG0581
PANA_0035	METABOLISM	Inorganic ion transport and metabolism	P	COG1117
PANA_0036	METABOLISM	Inorganic ion transport and metabolism	P	COG0704
PANA_0037	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_0038	METABOLISM	Carbohydrate transport and metabolism	G	COG1070
PANA_0039	METABOLISM	Carbohydrate transport and metabolism	G	COG2115
PANA_0040	METABOLISM	Carbohydrate transport and metabolism	G	COG4213
PANA_0041	METABOLISM	Carbohydrate transport and metabolism	G	COG1129
PANA_0042	METABOLISM	Carbohydrate transport and metabolism	G	COG4214
PANA_0043	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_0050	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1764
PANA_0056	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_0056	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840
PANA_0066	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1940
PANA_0066	METABOLISM	Carbohydrate transport and metabolism	G	COG1940
PANA_0067	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1484
PANA_0068	POORLY CHARACTERIZED	General function prediction only	R	COG5529
PANA_0069	NO COG	No COG association	NC	-
PANA_0070	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2818
PANA_0071	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2885
PANA_0072	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_0073	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609

PANA_0074	NO COG	No COG association	NC	-
PANA_0075	METABOLISM	Carbohydrate transport and metabolism	G	COG0524
PANA_0076	METABOLISM	Carbohydrate transport and metabolism	G	COG2271
PANA_0077	METABOLISM	Energy production and conversion	C	COG1052
PANA_0077	METABOLISM	Coenzyme transport and metabolism	H	COG1052
PANA_0077	POORLY CHARACTERIZED	General function prediction only	R	COG1052
PANA_0078	METABOLISM	Amino acid transport and metabolism	E	COG3977
PANA_0079	NO COG	No COG association	NC	-
PANA_0082	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0071
PANA_0083	POORLY CHARACTERIZED	General function prediction only	R	COG5645
PANA_0084	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0823
PANA_0086	POORLY CHARACTERIZED	General function prediction only	R	COG0561
PANA_0087	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0187
PANA_0088	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1195
PANA_0089	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0592
PANA_0090	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0593
PANA_0091	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0230
PANA_0092	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0594
PANA_0093	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0706
PANA_0094	POORLY CHARACTERIZED	General function prediction only	R	COG0486
PANA_0095	METABOLISM	Energy production and conversion	C	COG0778
PANA_0096	METABOLISM	Coenzyme transport and metabolism	H	COG2091
PANA_0097	POORLY CHARACTERIZED	General function prediction only	R	COG2252
PANA_0098	POORLY CHARACTERIZED	General function prediction only	R	COG0637
PANA_0099	POORLY CHARACTERIZED	Function unknown	S	COG3274
PANA_0100	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0752
PANA_0101	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0751
PANA_0104	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0625
PANA_0105	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG3058
PANA_0106	METABOLISM	Energy production and conversion	C	COG1526
PANA_0107	METABOLISM	Inorganic ion transport and metabolism	P	COG0605
PANA_0108	POORLY CHARACTERIZED	Function unknown	S	COG2990
PANA_0109	NO COG	No COG association	NC	-
PANA_0110	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3722
PANA_0111	METABOLISM	Carbohydrate transport and metabolism	G	COG0246
PANA_0112	METABOLISM	Carbohydrate transport and metabolism	G	COG2213
PANA_0113	NO COG	No COG association	NC	-
PANA_0114	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2944
PANA_0117	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0317
PANA_0117	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0317
PANA_0118	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1846
PANA_0120	POORLY CHARACTERIZED	General function prediction only	R	COG2194
PANA_0121	METABOLISM	Amino acid transport and metabolism	E	COG0747
PANA_0122	METABOLISM	Amino acid transport and metabolism	E	COG0601
PANA_0122	METABOLISM	Inorganic ion transport and metabolism	P	COG0601
PANA_0123	METABOLISM	Amino acid transport and metabolism	E	COG1173
PANA_0123	METABOLISM	Inorganic ion transport and metabolism	P	COG1173
PANA_0124	METABOLISM	Amino acid transport and metabolism	E	COG0444
PANA_0124	METABOLISM	Inorganic ion transport and metabolism	P	COG0444
PANA_0125	METABOLISM	Amino acid transport and metabolism	E	COG4608
PANA_0126	NO COG	No COG association	NC	-
PANA_0127	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG1192
PANA_0128	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1215
PANA_0129	NO COG	No COG association	NC	-
PANA_0130	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG3063
PANA_0130	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG3063
PANA_0132	METABOLISM	Carbohydrate transport and metabolism	G	COG3405
PANA_0133	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2200
PANA_0134	METABOLISM	Energy production and conversion	C	COG1301
PANA_0135	POORLY CHARACTERIZED	General function prediction only	R	COG0612
PANA_0136	METABOLISM	Lipid transport and metabolism	I	COG0657
PANA_0138	METABOLISM	Carbohydrate transport and metabolism	G	COG0524
PANA_0139	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2200
PANA_0140	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2982

PANA_0141	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0760
PANA_0142	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0210
PANA_0143	METABOLISM	Nucleotide transport and metabolism	F	COG0248
PANA_0143	METABOLISM	Inorganic ion transport and metabolism	P	COG0248
PANA_0144	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0513
PANA_0144	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0513
PANA_0144	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0513
PANA_0145	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG3118
PANA_0146	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1158
PANA_0147	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0472
PANA_0148	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3765
PANA_0149	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0381
PANA_0150	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0677
PANA_0151	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1088
PANA_0153	POORLY CHARACTERIZED	General function prediction only	R	COG0456
PANA_0154	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0399
PANA_0155	POORLY CHARACTERIZED	General function prediction only	R	COG2244
PANA_0156	NO COG	No COG association	NC	-
PANA_0157	NO COG	No COG association	NC	-
PANA_0158	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1922
PANA_0159	METABOLISM	Coenzyme transport and metabolism	H	COG3071
PANA_0160	METABOLISM	Coenzyme transport and metabolism	H	COG2959
PANA_0161	METABOLISM	Coenzyme transport and metabolism	H	COG1587
PANA_0162	METABOLISM	Coenzyme transport and metabolism	H	COG0181
PANA_0163	METABOLISM	Nucleotide transport and metabolism	F	COG3072
PANA_0164	NO COG	No COG association	NC	-
PANA_0165	METABOLISM	Amino acid transport and metabolism	E	COG0253
PANA_0166	POORLY CHARACTERIZED	Function unknown	S	COG3159
PANA_0167	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG4973
PANA_0168	POORLY CHARACTERIZED	General function prediction only	R	COG1011
PANA_0169	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0210
PANA_0170	METABOLISM	Inorganic ion transport and metabolism	P	COG0598
PANA_0171	POORLY CHARACTERIZED	General function prediction only	R	COG3180
PANA_0172	POORLY CHARACTERIZED	General function prediction only	R	COG2962
PANA_0175	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG2050
PANA_0176	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2829
PANA_0177	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0514
PANA_0178	METABOLISM	Amino acid transport and metabolism	E	COG1280
PANA_0179	METABOLISM	Amino acid transport and metabolism	E	COG1280
PANA_0180	METABOLISM	Lipid transport and metabolism	I	COG2267
PANA_0181	POORLY CHARACTERIZED	General function prediction only	R	COG0561
PANA_0183	METABOLISM	Carbohydrate transport and metabolism	G	COG0697
PANA_0183	METABOLISM	Amino acid transport and metabolism	E	COG0697
PANA_0183	POORLY CHARACTERIZED	General function prediction only	R	COG0697
PANA_0184	POORLY CHARACTERIZED	Function unknown	S	COG4807
PANA_0185	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_0186	METABOLISM	Amino acid transport and metabolism	E	COG0620
PANA_0187	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG0412
PANA_0188	METABOLISM	Nucleotide transport and metabolism	F	COG2820
PANA_0189	POORLY CHARACTERIZED	Function unknown	S	COG1322
PANA_0190	METABOLISM	Coenzyme transport and metabolism	H	COG2226
PANA_0191	POORLY CHARACTERIZED	Function unknown	S	COG3165
PANA_0192	POORLY CHARACTERIZED	General function prediction only	R	COG0661
PANA_0193	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1826
PANA_0194	NO COG	No COG association	NC	-
PANA_0195	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0805
PANA_0196	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0084
PANA_0197	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0250
PANA_0198	METABOLISM	Coenzyme transport and metabolism	H	COG0043
PANA_0199	METABOLISM	Coenzyme transport and metabolism	H	COG0543
PANA_0199	METABOLISM	Energy production and conversion	C	COG0543
PANA_0200	METABOLISM	Lipid transport and metabolism	I	COG0183
PANA_0201	METABOLISM	Lipid transport and metabolism	I	COG1250
PANA_0202	METABOLISM	Amino acid transport and metabolism	E	COG0006

PANA_0203	POORLY CHARACTERIZED	Function unknown	S	COG1739
PANA_0204	METABOLISM	Inorganic ion transport and metabolism	P	COG0168
PANA_0205	METABOLISM	Energy production and conversion	C	COG4635
PANA_0205	METABOLISM	Coenzyme transport and metabolism	H	COG4635
PANA_0206	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0812
PANA_0207	METABOLISM	Coenzyme transport and metabolism	H	COG0340
PANA_0208	METABOLISM	Coenzyme transport and metabolism	H	COG1072
PANA_0209	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0690
PANA_0210	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0250
PANA_0211	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0080
PANA_0212	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0081
PANA_0213	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0244
PANA_0214	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0222
PANA_0215	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0085
PANA_0216	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0086
PANA_0217	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1167
PANA_0217	METABOLISM	Amino acid transport and metabolism	E	COG1167
PANA_0218	METABOLISM	Coenzyme transport and metabolism	H	COG0352
PANA_0219	METABOLISM	Coenzyme transport and metabolism	H	COG0422
PANA_0220	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3160
PANA_0221	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2816
PANA_0222	METABOLISM	Coenzyme transport and metabolism	H	COG0407
PANA_0223	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1515
PANA_0224	POORLY CHARACTERIZED	Function unknown	S	COG3068
PANA_0225	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0776
PANA_0226	NO COG	No COG association	NC	-
PANA_0227	METABOLISM	Nucleotide transport and metabolism	F	COG0151
PANA_0228	METABOLISM	Nucleotide transport and metabolism	F	COG0138
PANA_0230	METABOLISM	Inorganic ion transport and metabolism	P	COG0614
PANA_0231	METABOLISM	Inorganic ion transport and metabolism	P	COG0725
PANA_0232	METABOLISM	Amino acid transport and metabolism	E	COG1897
PANA_0233	METABOLISM	Energy production and conversion	C	COG2225
PANA_0234	METABOLISM	Energy production and conversion	C	COG2224
PANA_0235	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG4579
PANA_0238	METABOLISM	Carbohydrate transport and metabolism	G	COG2211
PANA_0240	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1414
PANA_0241	METABOLISM	Amino acid transport and metabolism	E	COG1410
PANA_0242	NO COG	No COG association	NC	-
PANA_0243	NO COG	No COG association	NC	-
PANA_0244	NO COG	No COG association	NC	-
PANA_0245	NO COG	No COG association	NC	-
PANA_0246	METABOLISM	Carbohydrate transport and metabolism	G	COG0166
PANA_0247	METABOLISM	Amino acid transport and metabolism	E	COG0527
PANA_0248	POORLY CHARACTERIZED	General function prediction only	R	COG0385
PANA_0249	METABOLISM	Inorganic ion transport and metabolism	P	COG1283
PANA_0250	POORLY CHARACTERIZED	General function prediction only	R	COG1524
PANA_0251	POORLY CHARACTERIZED	Function unknown	S	COG3223
PANA_0252	METABOLISM	Coenzyme transport and metabolism	H	COG3161
PANA_0253	METABOLISM	Coenzyme transport and metabolism	H	COG0382
PANA_0254	METABOLISM	Lipid transport and metabolism	I	COG2937
PANA_0255	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0818
PANA_0256	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1974
PANA_0256	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1974
PANA_0257	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0534
PANA_0258	POORLY CHARACTERIZED	Function unknown	S	COG3237
PANA_0259	METABOLISM	Inorganic ion transport and metabolism	P	COG0735
PANA_0260	NO COG	No COG association	NC	-
PANA_0261	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0042
PANA_0262	NO COG	No COG association	NC	-
PANA_0263	METABOLISM	Energy production and conversion	C	COG0604
PANA_0263	POORLY CHARACTERIZED	General function prediction only	R	COG0604
PANA_0264	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0305
PANA_0265	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1522
PANA_0266	POORLY CHARACTERIZED	Function unknown	S	COG1284

PANA_0267	METABOLISM	Amino acid transport and metabolism	E	COG1448
PANA_0268	POORLY CHARACTERIZED	Function unknown	S	COG0432
PANA_0269	POORLY CHARACTERIZED	Function unknown	S	COG2315
PANA_0270	POORLY CHARACTERIZED	General function prediction only	R	COG1064
PANA_0272	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0178
PANA_0273	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0629
PANA_0275	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1335
PANA_0276	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_0278	POORLY CHARACTERIZED	General function prediction only	R	COG2244
PANA_0279	NO COG	No COG association	NC	-
PANA_0280	NO COG	No COG association	NC	-
PANA_0281	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_0282	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0789
PANA_0283	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0625
PANA_0284	POORLY CHARACTERIZED	General function prediction only	R	COG2252
PANA_0285	METABOLISM	Inorganic ion transport and metabolism	P	COG0025
PANA_0286	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_0287	POORLY CHARACTERIZED	General function prediction only	R	COG1380
PANA_0288	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1346
PANA_0289	NO COG	No COG association	NC	-
PANA_0290	POORLY CHARACTERIZED	General function prediction only	R	COG4147
PANA_0291	POORLY CHARACTERIZED	Function unknown	S	COG3162
PANA_0292	METABOLISM	Lipid transport and metabolism	I	COG0365
PANA_0293	METABOLISM	Energy production and conversion	C	COG1301
PANA_0294	POORLY CHARACTERIZED	General function prediction only	R	COG2103
PANA_0296	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_0296	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840
PANA_0297	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_0298	NO COG	No COG association	NC	-
PANA_0299	METABOLISM	Lipid transport and metabolism	I	COG1133
PANA_0300	NO COG	No COG association	NC	-
PANA_0301	METABOLISM	Amino acid transport and metabolism	E	COG2303
PANA_0302	METABOLISM	Energy production and conversion	C	COG2010
PANA_0303	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_0305	POORLY CHARACTERIZED	Function unknown	S	COG2764
PANA_0306	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3248
PANA_0307	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_0308	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_0308	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_0309	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG4977
PANA_0310	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_0311	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_0312	METABOLISM	Carbohydrate transport and metabolism	G	COG4993
PANA_0313	NO COG	No COG association	NC	-
PANA_0314	POORLY CHARACTERIZED	General function prediction only	R	COG0673
PANA_0315	METABOLISM	Carbohydrate transport and metabolism	G	COG2971
PANA_0316	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1940
PANA_0316	METABOLISM	Carbohydrate transport and metabolism	G	COG1940
PANA_0317	METABOLISM	Carbohydrate transport and metabolism	G	COG3839
PANA_0318	POORLY CHARACTERIZED	General function prediction only	R	COG0456
PANA_0319	METABOLISM	Carbohydrate transport and metabolism	G	COG1653
PANA_0320	METABOLISM	Carbohydrate transport and metabolism	G	COG1175
PANA_0321	METABOLISM	Carbohydrate transport and metabolism	G	COG0395
PANA_0324	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG5434
PANA_0325	METABOLISM	Energy production and conversion	C	COG0473
PANA_0325	METABOLISM	Amino acid transport and metabolism	E	COG0473
PANA_0326	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_0329	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0625
PANA_0330	METABOLISM	Carbohydrate transport and metabolism	G	COG0524
PANA_0331	METABOLISM	Carbohydrate transport and metabolism	G	COG3250
PANA_0333	METABOLISM	Carbohydrate transport and metabolism	G	COG4580
PANA_0334	METABOLISM	Carbohydrate transport and metabolism	G	COG1263
PANA_0335	METABOLISM	Carbohydrate transport and metabolism	G	COG1621
PANA_0336	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609

PANA_0344	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG3479
PANA_0345	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_0371	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1247
PANA_0374	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3659
PANA_0375	METABOLISM	Energy production and conversion	C	COG0667
PANA_0383	METABOLISM	Inorganic ion transport and metabolism	P	COG1918
PANA_0384	METABOLISM	Inorganic ion transport and metabolism	P	COG0370
PANA_0394	NO COG	No COG association	NC	-
PANA_0395	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG4232
PANA_0395	METABOLISM	Energy production and conversion	C	COG4232
PANA_0396	METABOLISM	Inorganic ion transport and metabolism	P	COG1324
PANA_0397	POORLY CHARACTERIZED	General function prediction only	R	COG2704
PANA_0398	METABOLISM	Amino acid transport and metabolism	E	COG1027
PANA_0399	POORLY CHARACTERIZED	General function prediction only	R	COG3030
PANA_0400	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0234
PANA_0401	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0459
PANA_0402	NO COG	No COG association	NC	-
PANA_0403	METABOLISM	Amino acid transport and metabolism	E	COG1509
PANA_0404	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0231
PANA_0406	POORLY CHARACTERIZED	Function unknown	S	COG5510
PANA_0407	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3040
PANA_0408	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2269
PANA_0409	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0265
PANA_0410	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0265
PANA_0411	POORLY CHARACTERIZED	Function unknown	S	COG3105
PANA_0412	POORLY CHARACTERIZED	General function prediction only	R	COG1485
PANA_0413	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0102
PANA_0414	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0103
PANA_0415	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0625
PANA_0416	POORLY CHARACTERIZED	General function prediction only	R	COG2969
PANA_0417	METABOLISM	Inorganic ion transport and metabolism	P	COG0369
PANA_0418	METABOLISM	Amino acid transport and metabolism	E	COG0069
PANA_0419	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_0420	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0744
PANA_0421	METABOLISM	Carbohydrate transport and metabolism	G	COG1925
PANA_0422	POORLY CHARACTERIZED	General function prediction only	R	COG1660
PANA_0423	METABOLISM	Carbohydrate transport and metabolism	G	COG1762
PANA_0423	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1762
PANA_0424	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1544
PANA_0425	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1508
PANA_0426	POORLY CHARACTERIZED	General function prediction only	R	COG1137
PANA_0427	POORLY CHARACTERIZED	Function unknown	S	COG1934
PANA_0428	POORLY CHARACTERIZED	Function unknown	S	COG3117
PANA_0429	POORLY CHARACTERIZED	General function prediction only	R	COG1778
PANA_0430	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0794
PANA_0431	METABOLISM	Inorganic ion transport and metabolism	P	COG0530
PANA_0432	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1127
PANA_0433	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG0767
PANA_0434	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1463
PANA_0435	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG2854
PANA_0436	POORLY CHARACTERIZED	General function prediction only	R	COG3113
PANA_0437	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG5007
PANA_0438	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0766
PANA_0439	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3423
PANA_0440	METABOLISM	Coenzyme transport and metabolism	H	COG0142
PANA_0441	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0261
PANA_0443	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0211
PANA_0444	POORLY CHARACTERIZED	General function prediction only	R	COG0536
PANA_0445	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG2050
PANA_0447	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2027
PANA_0448	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0782
PANA_0449	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1534
PANA_0450	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0293
PANA_0451	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0465

PANA_0452	METABOLISM	Coenzyme transport and metabolism	H	COG0294
PANA_0453	METABOLISM	Carbohydrate transport and metabolism	G	COG1109
PANA_0454	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1314
PANA_0455	POORLY CHARACTERIZED	Function unknown	S	COG0779
PANA_0456	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0195
PANA_0457	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0532
PANA_0458	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0858
PANA_0459	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0130
PANA_0460	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0184
PANA_0461	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1185
PANA_0462	POORLY CHARACTERIZED	General function prediction only	R	COG4785
PANA_0463	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0513
PANA_0463	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0513
PANA_0463	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0513
PANA_0464	METABOLISM	Energy production and conversion	C	COG2141
PANA_0465	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0826
PANA_0466	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0826
PANA_0467	METABOLISM	Lipid transport and metabolism	I	COG3154
PANA_0468	POORLY CHARACTERIZED	General function prediction only	R	COG3153
PANA_0469	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2827
PANA_0470	POORLY CHARACTERIZED	Function unknown	S	COG3787
PANA_0473	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0602
PANA_0474	METABOLISM	Nucleotide transport and metabolism	F	COG1328
PANA_0476	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_0477	METABOLISM	Energy production and conversion	C	COG0243
PANA_0478	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0768
PANA_0479	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0845
PANA_0480	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0841
PANA_0481	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0841
PANA_0482	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1538
PANA_0482	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1538
PANA_0483	METABOLISM	Lipid transport and metabolism	I	COG0318
PANA_0483	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG0318
PANA_0484	NO COG	No COG association	NC	-
PANA_0486	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0251
PANA_0487	METABOLISM	Nucleotide transport and metabolism	F	COG1781
PANA_0488	METABOLISM	Nucleotide transport and metabolism	F	COG0540
PANA_0489	METABOLISM	Amino acid transport and metabolism	E	COG0078
PANA_0490	POORLY CHARACTERIZED	Function unknown	S	COG3076
PANA_0491	METABOLISM	Nucleotide transport and metabolism	F	COG4445
PANA_0491	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG4445
PANA_0492	METABOLISM	Amino acid transport and metabolism	E	COG1246
PANA_0494	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0525
PANA_0496	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2927
PANA_0497	METABOLISM	Amino acid transport and metabolism	E	COG0260
PANA_0498	POORLY CHARACTERIZED	General function prediction only	R	COG0795
PANA_0499	POORLY CHARACTERIZED	General function prediction only	R	COG0795
PANA_0503	METABOLISM	Amino acid transport and metabolism	E	COG0624
PANA_0504	POORLY CHARACTERIZED	Function unknown	S	COG5426
PANA_0505	POORLY CHARACTERIZED	General function prediction only	R	COG1735
PANA_0506	METABOLISM	Carbohydrate transport and metabolism	G	COG1172
PANA_0507	METABOLISM	Carbohydrate transport and metabolism	G	COG1129
PANA_0508	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_0509	METABOLISM	Carbohydrate transport and metabolism	G	COG1879
PANA_0510	METABOLISM	Carbohydrate transport and metabolism	G	COG0524
PANA_0511	POORLY CHARACTERIZED	General function prediction only	R	COG0434
PANA_0512	POORLY CHARACTERIZED	Function unknown	S	COG2733
PANA_0514	POORLY CHARACTERIZED	General function prediction only	R	COG0433
PANA_0515	METABOLISM	Amino acid transport and metabolism	E	COG0747
PANA_0516	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG2095
PANA_0517	NO COG	No COG association	NC	-
PANA_0518	NO COG	No COG association	NC	-
PANA_0545	METABOLISM	Carbohydrate transport and metabolism	G	COG2731
PANA_0575	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207

PANA_0576	POORLY CHARACTERIZED	General function prediction only	R	COG2334
PANA_0577	METABOLISM	Amino acid transport and metabolism	E	COG0531
PANA_0578	METABOLISM	Amino acid transport and metabolism	E	COG0160
PANA_0580	POORLY CHARACTERIZED	General function prediction only	R	COG4225
PANA_0581	METABOLISM	Lipid transport and metabolism	I	COG1028
PANA_0581	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1028
PANA_0581	POORLY CHARACTERIZED	General function prediction only	R	COG1028
PANA_0582	METABOLISM	Carbohydrate transport and metabolism	G	COG3717
PANA_0592	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG4948
PANA_0592	POORLY CHARACTERIZED	General function prediction only	R	COG4948
PANA_0597	POORLY CHARACTERIZED	General function prediction only	R	COG3318
PANA_0599	POORLY CHARACTERIZED	Function unknown	S	COG5455
PANA_0600	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1670
PANA_0602	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1368
PANA_0603	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3774
PANA_0604	POORLY CHARACTERIZED	Function unknown	S	COG3391
PANA_0605	METABOLISM	Carbohydrate transport and metabolism	G	COG0406
PANA_0606	METABOLISM	Carbohydrate transport and metabolism	G	COG2271
PANA_0607	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_0608	METABOLISM	Nucleotide transport and metabolism	F	COG1957
PANA_0610	POORLY CHARACTERIZED	Function unknown	S	COG2606
PANA_0611	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2199
PANA_0612	NO COG	No COG association	NC	-
PANA_0613	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2813
PANA_0615	POORLY CHARACTERIZED	General function prediction only	R	COG0456
PANA_0616	POORLY CHARACTERIZED	General function prediction only	R	COG1011
PANA_0617	POORLY CHARACTERIZED	General function prediction only	R	COG3726
PANA_0618	METABOLISM	Amino acid transport and metabolism	E	COG0560
PANA_0619	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1066
PANA_0620	METABOLISM	Coenzyme transport and metabolism	H	COG3172
PANA_0628	METABOLISM	Energy production and conversion	C	COG0604
PANA_0628	POORLY CHARACTERIZED	General function prediction only	R	COG0604
PANA_0630	POORLY CHARACTERIZED	General function prediction only	R	COG0488
PANA_0631	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1349
PANA_0631	METABOLISM	Carbohydrate transport and metabolism	G	COG1349
PANA_0639	METABOLISM	Inorganic ion transport and metabolism	P	COG4773
PANA_0640	METABOLISM	Amino acid transport and metabolism	E	COG0410
PANA_0641	POORLY CHARACTERIZED	General function prediction only	R	COG4674
PANA_0642	METABOLISM	Amino acid transport and metabolism	E	COG4177
PANA_0643	METABOLISM	Amino acid transport and metabolism	E	COG0559
PANA_0644	METABOLISM	Amino acid transport and metabolism	E	COG0683
PANA_0645	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1802
PANA_0646	METABOLISM	Lipid transport and metabolism	I	COG4770
PANA_0647	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0741
PANA_0648	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2973
PANA_0649	POORLY CHARACTERIZED	Function unknown	S	COG1986
PANA_0650	METABOLISM	Carbohydrate transport and metabolism	G	COG0406
PANA_0651	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_0652	POORLY CHARACTERIZED	Function unknown	S	COG3045
PANA_0653	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_0653	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_0654	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0565
PANA_0655	METABOLISM	Amino acid transport and metabolism	E	COG0527
PANA_0656	METABOLISM	Amino acid transport and metabolism	E	COG0083
PANA_0657	METABOLISM	Amino acid transport and metabolism	E	COG0498
PANA_0658	POORLY CHARACTERIZED	Function unknown	S	COG3022
PANA_0659	METABOLISM	Carbohydrate transport and metabolism	G	COG0176
PANA_0660	METABOLISM	Coenzyme transport and metabolism	H	COG0521
PANA_0661	NO COG	No COG association	NC	-
PANA_0662	POORLY CHARACTERIZED	Function unknown	S	COG1584
PANA_0663	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0443
PANA_0664	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0484
PANA_0665	METABOLISM	Inorganic ion transport and metabolism	P	COG3004
PANA_0667	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0268

PANA_0668	METABOLISM	Coenzyme transport and metabolism	H	COG0196
PANA_0669	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0060
PANA_0670	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0597
PANA_0670	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0597
PANA_0671	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1047
PANA_0672	METABOLISM	Lipid transport and metabolism	I	COG0761
PANA_0672	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0761
PANA_0673	METABOLISM	Amino acid transport and metabolism	E	COG0289
PANA_0674	METABOLISM	Amino acid transport and metabolism	E	COG0505
PANA_0674	METABOLISM	Nucleotide transport and metabolism	F	COG0505
PANA_0675	METABOLISM	Amino acid transport and metabolism	E	COG0458
PANA_0675	METABOLISM	Nucleotide transport and metabolism	F	COG0458
PANA_0677	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0451
PANA_0677	METABOLISM	Carbohydrate transport and metabolism	G	COG0451
PANA_0678	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_0679	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3203
PANA_0680	METABOLISM	Coenzyme transport and metabolism	H	COG0262
PANA_0681	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0639
PANA_0682	METABOLISM	Inorganic ion transport and metabolism	P	COG2967
PANA_0683	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0030
PANA_0684	METABOLISM	Coenzyme transport and metabolism	H	COG1995
PANA_0685	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0760
PANA_0686	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1452
PANA_0687	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1076
PANA_0688	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0564
PANA_0689	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0553
PANA_0689	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0553
PANA_0690	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0417
PANA_0691	POORLY CHARACTERIZED	Function unknown	S	COG0586
PANA_0692	METABOLISM	Coenzyme transport and metabolism	H	COG3840
PANA_0693	METABOLISM	Inorganic ion transport and metabolism	P	COG1178
PANA_0694	METABOLISM	Coenzyme transport and metabolism	H	COG4143
PANA_0695	POORLY CHARACTERIZED	General function prediction only	R	COG4533
PANA_0696	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_0703	METABOLISM	Amino acid transport and metabolism	E	COG0066
PANA_0704	METABOLISM	Amino acid transport and metabolism	E	COG0065
PANA_0705	METABOLISM	Energy production and conversion	C	COG0473
PANA_0705	METABOLISM	Amino acid transport and metabolism	E	COG0473
PANA_0706	METABOLISM	Amino acid transport and metabolism	E	COG0119
PANA_0707	METABOLISM	Amino acid transport and metabolism	E	COG0028
PANA_0707	METABOLISM	Coenzyme transport and metabolism	H	COG0028
PANA_0708	METABOLISM	Amino acid transport and metabolism	E	COG0440
PANA_0709	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_0710	POORLY CHARACTERIZED	Function unknown	S	COG2001
PANA_0711	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0275
PANA_0712	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG3116
PANA_0713	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0768
PANA_0714	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0769
PANA_0715	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0770
PANA_0716	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0472
PANA_0717	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0771
PANA_0718	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0772
PANA_0719	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0707
PANA_0720	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0773
PANA_0721	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1181
PANA_0722	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1589
PANA_0723	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0849
PANA_0724	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0206
PANA_0725	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0774
PANA_0726	NO COG	No COG association	NC	-
PANA_0727	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0653
PANA_0728	METABOLISM	Nucleotide transport and metabolism	F	COG1051
PANA_0729	POORLY CHARACTERIZED	Function unknown	S	COG3024
PANA_0730	POORLY CHARACTERIZED	Function unknown	S	COG4582

PANA_0731	METABOLISM	Coenzyme transport and metabolism	H	COG0237
PANA_0732	METABOLISM	Nucleotide transport and metabolism	F	COG0516
PANA_0733	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1459
PANA_0733	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1459
PANA_0734	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG2804
PANA_0734	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG2804
PANA_0736	METABOLISM	Coenzyme transport and metabolism	H	COG0157
PANA_0737	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG3023
PANA_0738	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG3725
PANA_0739	POORLY CHARACTERIZED	General function prediction only	R	COG3940
PANA_0740	METABOLISM	Carbohydrate transport and metabolism	G	COG2211
PANA_0741	METABOLISM	Amino acid transport and metabolism	E	COG1113
PANA_0742	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2186
PANA_0743	METABOLISM	Energy production and conversion	C	COG2609
PANA_0744	METABOLISM	Energy production and conversion	C	COG0508
PANA_0745	METABOLISM	Energy production and conversion	C	COG1249
PANA_0747	METABOLISM	Energy production and conversion	C	COG0667
PANA_0748	METABOLISM	Energy production and conversion	C	COG1049
PANA_0749	POORLY CHARACTERIZED	Function unknown	S	COG3112
PANA_0751	POORLY CHARACTERIZED	Function unknown	S	COG3395
PANA_0755	METABOLISM	Amino acid transport and metabolism	E	COG1586
PANA_0756	METABOLISM	Amino acid transport and metabolism	E	COG0421
PANA_0757	NO COG	No COG association	NC	-
PANA_0758	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG2132
PANA_0759	METABOLISM	Amino acid transport and metabolism	E	COG0626
PANA_0760	METABOLISM	Nucleotide transport and metabolism	F	COG0634
PANA_0762	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1131
PANA_0763	METABOLISM	Carbohydrate transport and metabolism	G	COG1682
PANA_0763	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1682
PANA_0764	METABOLISM	Coenzyme transport and metabolism	H	COG0853
PANA_0765	METABOLISM	Coenzyme transport and metabolism	H	COG0414
PANA_0766	METABOLISM	Coenzyme transport and metabolism	H	COG0413
PANA_0767	METABOLISM	Coenzyme transport and metabolism	H	COG0801
PANA_0768	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0617
PANA_0769	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0008
PANA_0770	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1734
PANA_0771	POORLY CHARACTERIZED	General function prediction only	R	COG1489
PANA_0772	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1514
PANA_0773	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1643
PANA_0774	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0744
PANA_0775	METABOLISM	Inorganic ion transport and metabolism	P	COG4773
PANA_0776	METABOLISM	Inorganic ion transport and metabolism	P	COG1120
PANA_0776	METABOLISM	Coenzyme transport and metabolism	H	COG1120
PANA_0778	METABOLISM	Inorganic ion transport and metabolism	P	COG0609
PANA_0779	METABOLISM	Coenzyme transport and metabolism	H	COG0001
PANA_0780	POORLY CHARACTERIZED	Function unknown	S	COG0316
PANA_0782	METABOLISM	Inorganic ion transport and metabolism	P	COG0614
PANA_0783	METABOLISM	Nucleotide transport and metabolism	F	COG0775
PANA_0785	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0265
PANA_0786	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3835
PANA_0786	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3835
PANA_0788	NO COG	No COG association	NC	-
PANA_0789	METABOLISM	Amino acid transport and metabolism	E	COG2171
PANA_0790	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG2844
PANA_0791	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0024
PANA_0792	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0052
PANA_0793	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0264
PANA_0794	METABOLISM	Nucleotide transport and metabolism	F	COG0528
PANA_0795	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0233
PANA_0796	METABOLISM	Lipid transport and metabolism	I	COG0743
PANA_0797	METABOLISM	Lipid transport and metabolism	I	COG0020
PANA_0798	METABOLISM	Lipid transport and metabolism	I	COG0575
PANA_0799	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0750
PANA_0800	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG4775

PANA_0801	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2825
PANA_0802	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1044
PANA_0803	METABOLISM	Lipid transport and metabolism	I	COG0764
PANA_0805	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1043
PANA_0806	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0763
PANA_0807	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0164
PANA_0808	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0587
PANA_0809	METABOLISM	Lipid transport and metabolism	I	COG0825
PANA_0810	METABOLISM	Amino acid transport and metabolism	E	COG1982
PANA_0811	METABOLISM	Amino acid transport and metabolism	E	COG0346
PANA_0812	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0037
PANA_0814	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG4568
PANA_0816	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1186
PANA_0817	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0442
PANA_0818	POORLY CHARACTERIZED	Function unknown	S	COG1720
PANA_0819	NO COG	No COG association	NC	-
PANA_0820	METABOLISM	Inorganic ion transport and metabolism	P	COG1464
PANA_0821	METABOLISM	Inorganic ion transport and metabolism	P	COG2011
PANA_0822	METABOLISM	Inorganic ion transport and metabolism	P	COG1135
PANA_0823	METABOLISM	Amino acid transport and metabolism	E	COG0241
PANA_0824	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_0825	POORLY CHARACTERIZED	Function unknown	S	COG3021
PANA_0826	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0741
PANA_0827	POORLY CHARACTERIZED	General function prediction only	R	COG0491
PANA_0828	METABOLISM	Coenzyme transport and metabolism	H	COG2226
PANA_0829	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0328
PANA_0830	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0847
PANA_0839	METABOLISM	Nucleotide transport and metabolism	F	COG1953
PANA_0839	METABOLISM	Coenzyme transport and metabolism	H	COG1953
PANA_0840	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1802
PANA_0841	METABOLISM	Amino acid transport and metabolism	E	COG4126
PANA_0842	METABOLISM	Carbohydrate transport and metabolism	G	COG0726
PANA_0853	NO COG	No COG association	NC	-
PANA_0854	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_0856	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0154
PANA_0858	METABOLISM	Amino acid transport and metabolism	E	COG0405
PANA_0859	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1737
PANA_0860	METABOLISM	Amino acid transport and metabolism	E	COG0834
PANA_0860	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0834
PANA_0861	METABOLISM	Amino acid transport and metabolism	E	COG0765
PANA_0862	METABOLISM	Amino acid transport and metabolism	E	COG0765
PANA_0863	METABOLISM	Amino acid transport and metabolism	E	COG1126
PANA_0864	METABOLISM	Amino acid transport and metabolism	E	COG0075
PANA_0865	METABOLISM	Amino acid transport and metabolism	E	COG0624
PANA_0866	POORLY CHARACTERIZED	Function unknown	S	COG3195
PANA_0867	POORLY CHARACTERIZED	General function prediction only	R	COG2351
PANA_0868	METABOLISM	Inorganic ion transport and metabolism	P	COG0376
PANA_0869	POORLY CHARACTERIZED	General function prediction only	R	COG0388
PANA_0870	METABOLISM	Amino acid transport and metabolism	E	COG0436
PANA_0871	METABOLISM	Carbohydrate transport and metabolism	G	COG0235
PANA_0873	POORLY CHARACTERIZED	Function unknown	S	COG1791
PANA_0874	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0182
PANA_0875	POORLY CHARACTERIZED	General function prediction only	R	COG4857
PANA_0876	METABOLISM	Lipid transport and metabolism	I	COG1960
PANA_0877	METABOLISM	Carbohydrate transport and metabolism	G	COG0279
PANA_0878	POORLY CHARACTERIZED	General function prediction only	R	COG0121
PANA_0879	POORLY CHARACTERIZED	Function unknown	S	COG3034
PANA_0880	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0389
PANA_0881	METABOLISM	Carbohydrate transport and metabolism	G	COG1879
PANA_0882	METABOLISM	Carbohydrate transport and metabolism	G	COG1172
PANA_0883	METABOLISM	Carbohydrate transport and metabolism	G	COG1129
PANA_0887	METABOLISM	Amino acid transport and metabolism	E	COG2195
PANA_0888	METABOLISM	Nucleotide transport and metabolism	F	COG0503
PANA_0889	POORLY CHARACTERIZED	General function prediction only	R	COG1073

PANA_0890	NO COG	No COG association	NC	-
PANA_0891	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3203
PANA_0892	METABOLISM	Amino acid transport and metabolism	E	COG0263
PANA_0893	METABOLISM	Amino acid transport and metabolism	E	COG0014
PANA_0915	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1278
PANA_0916	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0835
PANA_0916	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0835
PANA_0918	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0513
PANA_0918	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0513
PANA_0918	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0513
PANA_0920	METABOLISM	Nucleotide transport and metabolism	F	COG0402
PANA_0920	POORLY CHARACTERIZED	General function prediction only	R	COG0402
PANA_0921	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_0922	METABOLISM	Inorganic ion transport and metabolism	P	COG2807
PANA_0923	METABOLISM	Energy production and conversion	C	COG1620
PANA_0924	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2186
PANA_0925	METABOLISM	Energy production and conversion	C	COG1304
PANA_0926	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1764
PANA_0934	NO COG	No COG association	NC	-
PANA_0939	NO COG	No COG association	NC	-
PANA_0940	NO COG	No COG association	NC	-
PANA_0941	POORLY CHARACTERIZED	General function prediction only	R	COG0446
PANA_0942	METABOLISM	Amino acid transport and metabolism	E	COG0345
PANA_0943	POORLY CHARACTERIZED	Function unknown	S	COG1671
PANA_0944	POORLY CHARACTERIZED	Function unknown	S	COG3416
PANA_0945	METABOLISM	Amino acid transport and metabolism	E	COG0703
PANA_0947	POORLY CHARACTERIZED	Function unknown	S	COG3123
PANA_0948	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2974
PANA_0949	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1940
PANA_0949	METABOLISM	Carbohydrate transport and metabolism	G	COG1940
PANA_0951	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0419
PANA_0952	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0420
PANA_0953	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_0953	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_0954	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG5002
PANA_0955	METABOLISM	Inorganic ion transport and metabolism	P	COG0226
PANA_0956	METABOLISM	Amino acid transport and metabolism	E	COG0620
PANA_0957	METABOLISM	Amino acid transport and metabolism	E	COG1114
PANA_0958	METABOLISM	Amino acid transport and metabolism	E	COG1113
PANA_0959	NO COG	No COG association	NC	-
PANA_0960	METABOLISM	Carbohydrate transport and metabolism	G	COG0483
PANA_0962	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0809
PANA_0963	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0343
PANA_0964	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1862
PANA_0965	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0342
PANA_0966	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0341
PANA_0967	NO COG	No COG association	NC	-
PANA_0968	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1327
PANA_0969	METABOLISM	Coenzyme transport and metabolism	H	COG0117
PANA_0970	METABOLISM	Coenzyme transport and metabolism	H	COG0054
PANA_0971	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0781
PANA_0972	METABOLISM	Coenzyme transport and metabolism	H	COG0611
PANA_0973	METABOLISM	Lipid transport and metabolism	I	COG1267
PANA_0975	METABOLISM	Energy production and conversion	C	COG0667
PANA_0976	METABOLISM	Coenzyme transport and metabolism	H	COG1154
PANA_0976	METABOLISM	Lipid transport and metabolism	I	COG1154
PANA_0977	METABOLISM	Coenzyme transport and metabolism	H	COG0142
PANA_0978	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1722
PANA_0979	METABOLISM	Coenzyme transport and metabolism	H	COG0301
PANA_0980	POORLY CHARACTERIZED	General function prediction only	R	COG0693
PANA_0981	METABOLISM	Coenzyme transport and metabolism	H	COG1893
PANA_0982	POORLY CHARACTERIZED	Function unknown	S	COG1666
PANA_0985	POORLY CHARACTERIZED	General function prediction only	R	COG1064
PANA_0986	METABOLISM	Carbohydrate transport and metabolism	G	COG2814

PANA_0987	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0109
PANA_0988	METABOLISM	Energy production and conversion	C	COG3125
PANA_0989	METABOLISM	Energy production and conversion	C	COG1845
PANA_0990	METABOLISM	Energy production and conversion	C	COG0843
PANA_0991	METABOLISM	Energy production and conversion	C	COG1622
PANA_0992	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_0993	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3056
PANA_0994	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0271
PANA_0995	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0544
PANA_0996	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0740
PANA_0996	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0740
PANA_0997	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1219
PANA_0998	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0466
PANA_0999	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0776
PANA_1000	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0760
PANA_1001	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1555
PANA_1002	POORLY CHARACTERIZED	General function prediction only	R	COG0824
PANA_1003	POORLY CHARACTERIZED	General function prediction only	R	COG0603
PANA_1004	METABOLISM	Amino acid transport and metabolism	E	COG0031
PANA_1005	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1522
PANA_1006	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1132
PANA_1007	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1132
PANA_1008	METABOLISM	Amino acid transport and metabolism	E	COG0347
PANA_1009	METABOLISM	Inorganic ion transport and metabolism	P	COG0004
PANA_1010	METABOLISM	Lipid transport and metabolism	I	COG1946
PANA_1011	POORLY CHARACTERIZED	Function unknown	S	COG3126
PANA_1012	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG3695
PANA_1013	NO COG	No COG association	NC	-
PANA_1014	NO COG	No COG association	NC	-
PANA_1016	METABOLISM	Inorganic ion transport and metabolism	P	COG1121
PANA_1017	METABOLISM	Inorganic ion transport and metabolism	P	COG1108
PANA_1018	METABOLISM	Inorganic ion transport and metabolism	P	COG0803
PANA_1019	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0257
PANA_1020	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0254
PANA_1021	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0841
PANA_1022	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0845
PANA_1023	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1309
PANA_1024	NO COG	No COG association	NC	-
PANA_1025	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG3923
PANA_1026	POORLY CHARACTERIZED	Function unknown	S	COG2832
PANA_1027	METABOLISM	Nucleotide transport and metabolism	F	COG0503
PANA_1029	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2812
PANA_1030	POORLY CHARACTERIZED	Function unknown	S	COG0718
PANA_1031	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0353
PANA_1033	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0326
PANA_1034	METABOLISM	Nucleotide transport and metabolism	F	COG0563
PANA_1035	METABOLISM	Carbohydrate transport and metabolism	G	COG0524
PANA_1036	METABOLISM	Inorganic ion transport and metabolism	P	COG4651
PANA_1037	METABOLISM	Nucleotide transport and metabolism	F	COG0737
PANA_1039	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_1040	POORLY CHARACTERIZED	Function unknown	S	COG2606
PANA_1041	POORLY CHARACTERIZED	Function unknown	S	COG3735
PANA_1042	METABOLISM	Inorganic ion transport and metabolism	P	COG2217
PANA_1044	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0789
PANA_1045	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1585
PANA_1045	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1585
PANA_1046	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0330
PANA_1047	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG3118
PANA_1048	POORLY CHARACTERIZED	General function prediction only	R	COG0300
PANA_1049	METABOLISM	Amino acid transport and metabolism	E	COG2755
PANA_1050	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG4181
PANA_1051	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG3127
PANA_1052	METABOLISM	Nucleotide transport and metabolism	F	COG0026
PANA_1053	METABOLISM	Nucleotide transport and metabolism	F	COG0041

PANA_1055	POORLY CHARACTERIZED	Function unknown	S	COG2908
PANA_1056	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0652
PANA_1057	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0215
PANA_1058	POORLY CHARACTERIZED	General function prediction only	R	COG4637
PANA_1059	POORLY CHARACTERIZED	Function unknown	S	COG2501
PANA_1060	METABOLISM	Coenzyme transport and metabolism	H	COG0190
PANA_1061	NO COG	No COG association	NC	-
PANA_1097	POORLY CHARACTERIZED	Function unknown	S	COG3619
PANA_1100	NO COG	No COG association	NC	-
PANA_1102	POORLY CHARACTERIZED	General function prediction only	R	COG0628
PANA_1105	METABOLISM	Amino acid transport and metabolism	E	COG0814
PANA_1106	METABOLISM	Inorganic ion transport and metabolism	P	COG0614
PANA_1107	POORLY CHARACTERIZED	Function unknown	S	COG5455
PANA_1108	POORLY CHARACTERIZED	General function prediction only	R	COG0561
PANA_1109	METABOLISM	Energy production and conversion	C	COG1454
PANA_1110	METABOLISM	Carbohydrate transport and metabolism	G	COG2610
PANA_1110	METABOLISM	Amino acid transport and metabolism	E	COG2610
PANA_1113	NO COG	No COG association	NC	-
PANA_1114	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1278
PANA_1115	NO COG	No COG association	NC	-
PANA_1116	POORLY CHARACTERIZED	General function prediction only	R	COG0388
PANA_1117	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1826
PANA_1118	METABOLISM	Coenzyme transport and metabolism	H	COG0320
PANA_1119	METABOLISM	Coenzyme transport and metabolism	H	COG0321
PANA_1120	POORLY CHARACTERIZED	Function unknown	S	COG2921
PANA_1121	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1686
PANA_1122	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0797
PANA_1123	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0772
PANA_1124	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0768
PANA_1125	POORLY CHARACTERIZED	Function unknown	S	COG1576
PANA_1126	POORLY CHARACTERIZED	Function unknown	S	COG0799
PANA_1127	METABOLISM	Coenzyme transport and metabolism	H	COG1057
PANA_1128	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1466
PANA_1129	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2980
PANA_1130	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0495
PANA_1131	NO COG	No COG association	NC	-
PANA_1132	METABOLISM	Amino acid transport and metabolism	E	COG1126
PANA_1133	METABOLISM	Amino acid transport and metabolism	E	COG0765
PANA_1134	METABOLISM	Amino acid transport and metabolism	E	COG0765
PANA_1135	METABOLISM	Amino acid transport and metabolism	E	COG0834
PANA_1135	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0834
PANA_1136	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0815
PANA_1137	METABOLISM	Inorganic ion transport and metabolism	P	COG4535
PANA_1139	POORLY CHARACTERIZED	General function prediction only	R	COG0319
PANA_1140	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1702
PANA_1141	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0621
PANA_1142	METABOLISM	Coenzyme transport and metabolism	H	COG0654
PANA_1142	METABOLISM	Energy production and conversion	C	COG0654
PANA_1143	METABOLISM	Amino acid transport and metabolism	E	COG0367
PANA_1144	METABOLISM	Carbohydrate transport and metabolism	G	COG0647
PANA_1145	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1940
PANA_1145	METABOLISM	Carbohydrate transport and metabolism	G	COG1940
PANA_1146	METABOLISM	Carbohydrate transport and metabolism	G	COG1820
PANA_1147	METABOLISM	Carbohydrate transport and metabolism	G	COG0363
PANA_1148	METABOLISM	Carbohydrate transport and metabolism	G	COG1263
PANA_1149	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0008
PANA_1151	METABOLISM	Inorganic ion transport and metabolism	P	COG0735
PANA_1152	METABOLISM	Energy production and conversion	C	COG0716
PANA_1153	NO COG	No COG association	NC	-
PANA_1154	POORLY CHARACTERIZED	General function prediction only	R	COG0596
PANA_1155	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG3057
PANA_1156	METABOLISM	Carbohydrate transport and metabolism	G	COG0033
PANA_1157	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_1157	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745

PANA_1158	NO COG	No COG association	NC	-
PANA_1159	POORLY CHARACTERIZED	Function unknown	S	COG3272
PANA_1160	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0415
PANA_1161	POORLY CHARACTERIZED	Function unknown	S	COG0327
PANA_1162	METABOLISM	Amino acid transport and metabolism	E	COG2049
PANA_1163	METABOLISM	Amino acid transport and metabolism	E	COG1984
PANA_1164	POORLY CHARACTERIZED	General function prediction only	R	COG1540
PANA_1165	POORLY CHARACTERIZED	Function unknown	S	COG3819
PANA_1166	POORLY CHARACTERIZED	Function unknown	S	COG3817
PANA_1167	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG2039
PANA_1168	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0266
PANA_1169	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1670
PANA_1170	METABOLISM	Energy production and conversion	C	COG0372
PANA_1171	METABOLISM	Energy production and conversion	C	COG2009
PANA_1172	METABOLISM	Energy production and conversion	C	COG2142
PANA_1173	METABOLISM	Energy production and conversion	C	COG1053
PANA_1174	METABOLISM	Energy production and conversion	C	COG0479
PANA_1175	METABOLISM	Energy production and conversion	C	COG0567
PANA_1176	METABOLISM	Energy production and conversion	C	COG0508
PANA_1177	METABOLISM	Energy production and conversion	C	COG0045
PANA_1179	METABOLISM	Energy production and conversion	C	COG0074
PANA_1180	METABOLISM	Energy production and conversion	C	COG1271
PANA_1181	METABOLISM	Energy production and conversion	C	COG1294
PANA_1183	POORLY CHARACTERIZED	General function prediction only	R	COG0824
PANA_1184	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0811
PANA_1185	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0848
PANA_1186	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3064
PANA_1187	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0823
PANA_1188	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2885
PANA_1189	POORLY CHARACTERIZED	Function unknown	S	COG1729
PANA_1191	METABOLISM	Nucleotide transport and metabolism	F	COG2169
PANA_1192	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0350
PANA_1193	METABOLISM	Coenzyme transport and metabolism	H	COG0379
PANA_1194	METABOLISM	Coenzyme transport and metabolism	H	COG3201
PANA_1195	METABOLISM	Inorganic ion transport and metabolism	P	COG1230
PANA_1196	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_1197	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2197
PANA_1197	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2197
PANA_1198	NO COG	No COG association	NC	-
PANA_1199	METABOLISM	Amino acid transport and metabolism	E	COG0722
PANA_1200	METABOLISM	Carbohydrate transport and metabolism	G	COG0588
PANA_1201	METABOLISM	Carbohydrate transport and metabolism	G	COG2017
PANA_1202	METABOLISM	Carbohydrate transport and metabolism	G	COG0153
PANA_1203	METABOLISM	Energy production and conversion	C	COG1085
PANA_1204	METABOLISM	Inorganic ion transport and metabolism	P	COG1119
PANA_1205	POORLY CHARACTERIZED	General function prediction only	R	COG2005
PANA_1207	METABOLISM	Inorganic ion transport and metabolism	P	COG0725
PANA_1208	METABOLISM	Inorganic ion transport and metabolism	P	COG4149
PANA_1209	METABOLISM	Inorganic ion transport and metabolism	P	COG4148
PANA_1210	POORLY CHARACTERIZED	General function prediction only	R	COG0561
PANA_1211	METABOLISM	Carbohydrate transport and metabolism	G	COG2706
PANA_1213	METABOLISM	Coenzyme transport and metabolism	H	COG0161
PANA_1214	METABOLISM	Coenzyme transport and metabolism	H	COG0502
PANA_1215	METABOLISM	Coenzyme transport and metabolism	H	COG0156
PANA_1216	METABOLISM	Coenzyme transport and metabolism	H	COG2226
PANA_1217	METABOLISM	Coenzyme transport and metabolism	H	COG0132
PANA_1218	METABOLISM	Amino acid transport and metabolism	E	COG0410
PANA_1219	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0556
PANA_1220	POORLY CHARACTERIZED	Function unknown	S	COG0391
PANA_1221	METABOLISM	Coenzyme transport and metabolism	H	COG2896
PANA_1222	METABOLISM	Coenzyme transport and metabolism	H	COG0521
PANA_1223	METABOLISM	Coenzyme transport and metabolism	H	COG0315
PANA_1224	METABOLISM	Coenzyme transport and metabolism	H	COG1977
PANA_1225	METABOLISM	Coenzyme transport and metabolism	H	COG0314

PANA_1226	NO COG	No COG association	NC	-
PANA_1227	POORLY CHARACTERIZED	General function prediction only	R	COG0670
PANA_1228	POORLY CHARACTERIZED	Function unknown	S	COG0392
PANA_1229	METABOLISM	Lipid transport and metabolism	I	COG1502
PANA_1230	POORLY CHARACTERIZED	General function prediction only	R	COG3568
PANA_1231	NO COG	No COG association	NC	-
PANA_1232	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0513
PANA_1232	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0513
PANA_1232	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0513
PANA_1233	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0042
PANA_1236	METABOLISM	Inorganic ion transport and metabolism	P	COG0306
PANA_1238	POORLY CHARACTERIZED	Function unknown	S	COG1376
PANA_1239	METABOLISM	Carbohydrate transport and metabolism	G	COG3387
PANA_1241	POORLY CHARACTERIZED	Function unknown	S	COG0586
PANA_1242	NO COG	No COG association	NC	-
PANA_1243	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1686
PANA_1244	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3264
PANA_1245	METABOLISM	Energy production and conversion	C	COG0277
PANA_1246	METABOLISM	Carbohydrate transport and metabolism	G	COG1472
PANA_1247	NO COG	No COG association	NC	-
PANA_1248	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_1248	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840
PANA_1249	METABOLISM	Lipid transport and metabolism	I	COG1443
PANA_1250	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1732
PANA_1251	METABOLISM	Amino acid transport and metabolism	E	COG1174
PANA_1252	METABOLISM	Amino acid transport and metabolism	E	COG1125
PANA_1253	METABOLISM	Amino acid transport and metabolism	E	COG1174
PANA_1255	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0789
PANA_1256	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1396
PANA_1257	POORLY CHARACTERIZED	General function prediction only	R	COG0456
PANA_1258	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1199
PANA_1258	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1199
PANA_1259	METABOLISM	Amino acid transport and metabolism	E	COG0547
PANA_1260	POORLY CHARACTERIZED	General function prediction only	R	COG1473
PANA_1264	METABOLISM	Amino acid transport and metabolism	E	COG2987
PANA_1265	METABOLISM	Amino acid transport and metabolism	E	COG2986
PANA_1266	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2188
PANA_1267	POORLY CHARACTERIZED	Function unknown	S	COG3758
PANA_1268	METABOLISM	Nucleotide transport and metabolism	F	COG0402
PANA_1268	POORLY CHARACTERIZED	General function prediction only	R	COG0402
PANA_1269	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1228
PANA_1270	METABOLISM	Amino acid transport and metabolism	E	COG3741
PANA_1272	METABOLISM	Amino acid transport and metabolism	E	COG0834
PANA_1272	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0834
PANA_1273	METABOLISM	Amino acid transport and metabolism	E	COG4215
PANA_1274	METABOLISM	Amino acid transport and metabolism	E	COG4160
PANA_1275	METABOLISM	Amino acid transport and metabolism	E	COG4598
PANA_1276	METABOLISM	Amino acid transport and metabolism	E	COG2986
PANA_1277	POORLY CHARACTERIZED	General function prediction only	R	COG1853
PANA_1279	POORLY CHARACTERIZED	General function prediction only	R	COG3129
PANA_1280	METABOLISM	Amino acid transport and metabolism	E	COG1126
PANA_1282	METABOLISM	Inorganic ion transport and metabolism	P	COG0783
PANA_1283	POORLY CHARACTERIZED	General function prediction only	R	COG5006
PANA_1284	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3637
PANA_1286	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1321
PANA_1287	METABOLISM	Inorganic ion transport and metabolism	P	COG1055
PANA_1288	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1566
PANA_1289	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_1290	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1538
PANA_1290	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1538
PANA_1291	POORLY CHARACTERIZED	General function prediction only	R	COG2916
PANA_1292	POORLY CHARACTERIZED	General function prediction only	R	COG0488
PANA_1293	METABOLISM	Coenzyme transport and metabolism	H	COG0476
PANA_1294	METABOLISM	Coenzyme transport and metabolism	H	COG0303

PANA_1295	METABOLISM	Amino acid transport and metabolism	E	COG1446
PANA_1296	POORLY CHARACTERIZED	General function prediction only	R	COG1123
PANA_1297	METABOLISM	Amino acid transport and metabolism	E	COG0747
PANA_1298	METABOLISM	Amino acid transport and metabolism	E	COG0601
PANA_1298	METABOLISM	Inorganic ion transport and metabolism	P	COG0601
PANA_1299	METABOLISM	Amino acid transport and metabolism	E	COG1173
PANA_1299	METABOLISM	Inorganic ion transport and metabolism	P	COG1173
PANA_1302	METABOLISM	Inorganic ion transport and metabolism	P	COG0288
PANA_1303	METABOLISM	Carbohydrate transport and metabolism	G	COG2133
PANA_1304	METABOLISM	Amino acid transport and metabolism	E	COG0560
PANA_1305	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1686
PANA_1306	NO COG	No COG association	NC	-
PANA_1307	POORLY CHARACTERIZED	General function prediction only	R	COG2985
PANA_1308	NO COG	No COG association	NC	-
PANA_1309	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0695
PANA_1310	NO COG	No COG association	NC	-
PANA_1311	METABOLISM	Energy production and conversion	C	COG0778
PANA_1312	METABOLISM	Coenzyme transport and metabolism	H	COG0189
PANA_1312	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0189
PANA_1313	NO COG	No COG association	NC	-
PANA_1314	METABOLISM	Amino acid transport and metabolism	E	COG0687
PANA_1315	METABOLISM	Amino acid transport and metabolism	E	COG3842
PANA_1316	METABOLISM	Amino acid transport and metabolism	E	COG1176
PANA_1317	METABOLISM	Amino acid transport and metabolism	E	COG1177
PANA_1318	NO COG	No COG association	NC	-
PANA_1319	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2265
PANA_1320	METABOLISM	Amino acid transport and metabolism	E	COG0834
PANA_1320	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0834
PANA_1321	METABOLISM	Amino acid transport and metabolism	E	COG4160
PANA_1322	METABOLISM	Amino acid transport and metabolism	E	COG4215
PANA_1323	METABOLISM	Amino acid transport and metabolism	E	COG0834
PANA_1323	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0834
PANA_1324	METABOLISM	Amino acid transport and metabolism	E	COG4161
PANA_1325	NO COG	No COG association	NC	-
PANA_1327	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG3023
PANA_1328	METABOLISM	Amino acid transport and metabolism	E	COG2008
PANA_1329	METABOLISM	Amino acid transport and metabolism	E	COG0028
PANA_1329	METABOLISM	Coenzyme transport and metabolism	H	COG0028
PANA_1330	POORLY CHARACTERIZED	Function unknown	S	COG2431
PANA_1331	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG3593
PANA_1332	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1278
PANA_1333	POORLY CHARACTERIZED	Function unknown	S	COG2127
PANA_1334	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0542
PANA_1335	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0361
PANA_1336	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG2360
PANA_1337	METABOLISM	Energy production and conversion	C	COG4987
PANA_1337	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG4987
PANA_1338	METABOLISM	Energy production and conversion	C	COG4988
PANA_1338	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG4988
PANA_1339	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0492
PANA_1340	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1522
PANA_1341	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG1674
PANA_1342	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2834
PANA_1343	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2256
PANA_1344	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0172
PANA_1345	METABOLISM	Carbohydrate transport and metabolism	G	COG0738
PANA_1346	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1180
PANA_1347	METABOLISM	Energy production and conversion	C	COG1882
PANA_1348	METABOLISM	Inorganic ion transport and metabolism	P	COG2116
PANA_1349	POORLY CHARACTERIZED	Function unknown	S	COG1944
PANA_1350	METABOLISM	Coenzyme transport and metabolism	H	COG1932
PANA_1350	METABOLISM	Amino acid transport and metabolism	E	COG1932
PANA_1351	METABOLISM	Amino acid transport and metabolism	E	COG0128
PANA_1352	METABOLISM	Nucleotide transport and metabolism	F	COG0283

PANA_1353	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0539
PANA_1354	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0776
PANA_1356	POORLY CHARACTERIZED	General function prediction only	R	COG2333
PANA_1357	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1132
PANA_1358	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1663
PANA_1359	POORLY CHARACTERIZED	Function unknown	S	COG3214
PANA_1360	POORLY CHARACTERIZED	Function unknown	S	COG2835
PANA_1361	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1212
PANA_1362	POORLY CHARACTERIZED	Function unknown	S	COG1434
PANA_1363	METABOLISM	Coenzyme transport and metabolism	H	COG2226
PANA_1365	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG3006
PANA_1364	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG3095
PANA_1366	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG3096
PANA_1367	POORLY CHARACTERIZED	Function unknown	S	COG2989
PANA_1368	POORLY CHARACTERIZED	Function unknown	S	COG3108
PANA_1369	POORLY CHARACTERIZED	General function prediction only	R	COG0491
PANA_1370	METABOLISM	Amino acid transport and metabolism	E	COG1448
PANA_1371	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3203
PANA_1372	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0017
PANA_1373	METABOLISM	Coenzyme transport and metabolism	H	COG1488
PANA_1374	METABOLISM	Amino acid transport and metabolism	E	COG0308
PANA_1375	METABOLISM	Inorganic ion transport and metabolism	P	COG1116
PANA_1376	METABOLISM	Inorganic ion transport and metabolism	P	COG0600
PANA_1377	METABOLISM	Inorganic ion transport and metabolism	P	COG0715
PANA_1378	POORLY CHARACTERIZED	General function prediction only	R	COG0431
PANA_1379	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0463
PANA_1380	POORLY CHARACTERIZED	General function prediction only	R	COG2244
PANA_1381	METABOLISM	Nucleotide transport and metabolism	F	COG0167
PANA_1382	NO COG	No COG association	NC	-
PANA_1383	POORLY CHARACTERIZED	General function prediction only	R	COG3217
PANA_1384	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0116
PANA_1385	POORLY CHARACTERIZED	General function prediction only	R	COG0488
PANA_1386	POORLY CHARACTERIZED	Function unknown	S	COG2995
PANA_1387	POORLY CHARACTERIZED	General function prediction only	R	COG3008
PANA_1388	POORLY CHARACTERIZED	Function unknown	S	COG3009
PANA_1389	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG3130
PANA_1390	METABOLISM	Lipid transport and metabolism	I	COG0764
PANA_1391	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1067
PANA_1392	POORLY CHARACTERIZED	Function unknown	S	COG3120
PANA_1393	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2885
PANA_1394	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG5404
PANA_1395	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3070
PANA_1396	POORLY CHARACTERIZED	Function unknown	S	COG1289
PANA_1397	POORLY CHARACTERIZED	Function unknown	S	COG3304
PANA_1398	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0210
PANA_1399	METABOLISM	Carbohydrate transport and metabolism	G	COG1803
PANA_1400	POORLY CHARACTERIZED	Function unknown	S	COG3110
PANA_1401	POORLY CHARACTERIZED	General function prediction only	R	COG1832
PANA_1402	POORLY CHARACTERIZED	Function unknown	S	COG3785
PANA_1403	POORLY CHARACTERIZED	General function prediction only	R	COG1092
PANA_1404	METABOLISM	Energy production and conversion	C	COG1254
PANA_1405	METABOLISM	Inorganic ion transport and metabolism	P	COG2920
PANA_1406	POORLY CHARACTERIZED	General function prediction only	R	COG0670
PANA_1407	NO COG	No COG association	NC	-
PANA_1408	METABOLISM	Energy production and conversion	C	COG1012
PANA_1409	NO COG	No COG association	NC	-
PANA_1410	POORLY CHARACTERIZED	General function prediction only	R	COG0655
PANA_1411	NO COG	No COG association	NC	-
PANA_1412	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1309
PANA_1413	POORLY CHARACTERIZED	General function prediction only	R	COG1988
PANA_1414	METABOLISM	Amino acid transport and metabolism	E	COG1113
PANA_1415	POORLY CHARACTERIZED	Function unknown	S	COG3921
PANA_1416	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2771
PANA_1417	NO COG	No COG association	NC	-

PANA_1418	NO COG	No COG association	NC	-
PANA_1419	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2197
PANA_1419	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2197
PANA_1420	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0322
PANA_1421	METABOLISM	Lipid transport and metabolism	I	COG0558
PANA_1426	METABOLISM	Inorganic ion transport and metabolism	P	COG0659
PANA_1427	NO COG	No COG association	NC	-
PANA_1429	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_1430	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1846
PANA_1431	NO COG	No COG association	NC	-
PANA_1432	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1566
PANA_1433	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1538
PANA_1433	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1538
PANA_1436	NO COG	No COG association	NC	-
PANA_1437	METABOLISM	Lipid transport and metabolism	I	COG1502
PANA_1438	METABOLISM	Carbohydrate transport and metabolism	G	COG3594
PANA_1439	METABOLISM	Inorganic ion transport and metabolism	P	COG3131
PANA_1440	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2943
PANA_1441	POORLY CHARACTERIZED	General function prediction only	R	COG5645
PANA_1442	NO COG	No COG association	NC	-
PANA_1447	POORLY CHARACTERIZED	General function prediction only	R	COG1054
PANA_1448	POORLY CHARACTERIZED	Function unknown	S	COG2353
PANA_1449	METABOLISM	Energy production and conversion	C	COG3038
PANA_1450	METABOLISM	Amino acid transport and metabolism	E	COG0665
PANA_1451	NO COG	No COG association	NC	-
PANA_1453	NO COG	No COG association	NC	-
PANA_1454	METABOLISM	Nucleotide transport and metabolism	F	COG0418
PANA_1455	NO COG	No COG association	NC	-
PANA_1456	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1670
PANA_1457	POORLY CHARACTERIZED	Function unknown	S	COG3132
PANA_1458	POORLY CHARACTERIZED	General function prediction only	R	COG0673
PANA_1459	POORLY CHARACTERIZED	General function prediction only	R	COG0728
PANA_1461	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG3418
PANA_1461	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG3418
PANA_1461	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG3418
PANA_1462	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2747
PANA_1462	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG2747
PANA_1462	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG2747
PANA_1463	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1261
PANA_1463	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1261
PANA_1464	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1815
PANA_1465	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1558
PANA_1466	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1843
PANA_1467	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1749
PANA_1468	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG4787
PANA_1469	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG4786
PANA_1470	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG2063
PANA_1471	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1706
PANA_1472	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1705
PANA_1472	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1705
PANA_1473	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1256
PANA_1474	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1344
PANA_1475	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1530
PANA_1476	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0564
PANA_1477	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0424
PANA_1478	POORLY CHARACTERIZED	General function prediction only	R	COG1399
PANA_1479	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0333
PANA_1480	METABOLISM	Lipid transport and metabolism	I	COG0416
PANA_1481	METABOLISM	Lipid transport and metabolism	I	COG0332
PANA_1482	METABOLISM	Lipid transport and metabolism	I	COG0331
PANA_1483	METABOLISM	Lipid transport and metabolism	I	COG1028
PANA_1483	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1028
PANA_1483	POORLY CHARACTERIZED	General function prediction only	R	COG1028
PANA_1484	METABOLISM	Lipid transport and metabolism	I	COG0236

PANA_1484	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG0236
PANA_1485	METABOLISM	Lipid transport and metabolism	I	COG0304
PANA_1485	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG0304
PANA_1486	METABOLISM	Amino acid transport and metabolism	E	COG0115
PANA_1486	METABOLISM	Coenzyme transport and metabolism	H	COG0115
PANA_1487	POORLY CHARACTERIZED	General function prediction only	R	COG1559
PANA_1488	METABOLISM	Nucleotide transport and metabolism	F	COG0125
PANA_1489	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0470
PANA_1490	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0084
PANA_1491	METABOLISM	Carbohydrate transport and metabolism	G	COG1263
PANA_1493	METABOLISM	Nucleotide transport and metabolism	F	COG0537
PANA_1493	METABOLISM	Carbohydrate transport and metabolism	G	COG0537
PANA_1493	POORLY CHARACTERIZED	General function prediction only	R	COG0537
PANA_1494	POORLY CHARACTERIZED	General function prediction only	R	COG5633
PANA_1495	POORLY CHARACTERIZED	General function prediction only	R	COG3417
PANA_1496	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0510
PANA_1497	METABOLISM	Carbohydrate transport and metabolism	G	COG1472
PANA_1498	POORLY CHARACTERIZED	General function prediction only	R	COG3150
PANA_1499	METABOLISM	Energy production and conversion	C	COG1252
PANA_1500	POORLY CHARACTERIZED	Function unknown	S	COG3134
PANA_1502	POORLY CHARACTERIZED	Function unknown	S	COG2128
PANA_1503	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1197
PANA_1503	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1197
PANA_1505	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG4591
PANA_1506	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1136
PANA_1507	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG4591
PANA_1508	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0846
PANA_1509	METABOLISM	Amino acid transport and metabolism	E	COG2195
PANA_1510	POORLY CHARACTERIZED	Function unknown	S	COG2850
PANA_1511	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_1512	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_1512	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_1513	METABOLISM	Nucleotide transport and metabolism	F	COG0015
PANA_1514	POORLY CHARACTERIZED	General function prediction only	R	COG2915
PANA_1515	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0482
PANA_1516	METABOLISM	Nucleotide transport and metabolism	F	COG1051
PANA_1517	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1187
PANA_1518	METABOLISM	Energy production and conversion	C	COG0538
PANA_1527	METABOLISM	Coenzyme transport and metabolism	H	COG0654
PANA_1527	METABOLISM	Energy production and conversion	C	COG0654
PANA_1528	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1309
PANA_1529	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1538
PANA_1529	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1538
PANA_1530	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0841
PANA_1531	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0845
PANA_1533	NO COG	No COG association	NC	-
PANA_1539	NO COG	No COG association	NC	-
PANA_1543	POORLY CHARACTERIZED	General function prediction only	R	COG0456
PANA_1544	POORLY CHARACTERIZED	General function prediction only	R	COG0491
PANA_1545	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1619
PANA_1546	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0741
PANA_1547	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2390
PANA_1548	METABOLISM	Carbohydrate transport and metabolism	G	COG3959
PANA_1549	METABOLISM	Carbohydrate transport and metabolism	G	COG3958
PANA_1550	METABOLISM	Energy production and conversion	C	COG0554
PANA_1552	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG5581
PANA_1553	POORLY CHARACTERIZED	Function unknown	S	COG2261
PANA_1554	METABOLISM	Carbohydrate transport and metabolism	G	COG1626
PANA_1555	METABOLISM	Coenzyme transport and metabolism	H	COG0111
PANA_1555	METABOLISM	Amino acid transport and metabolism	E	COG0111
PANA_1556	METABOLISM	Amino acid transport and metabolism	E	COG1387
PANA_1556	POORLY CHARACTERIZED	General function prediction only	R	COG1387
PANA_1557	POORLY CHARACTERIZED	General function prediction only	R	COG3381
PANA_1558	METABOLISM	Energy production and conversion	C	COG1804

PANA_1559	METABOLISM	Carbohydrate transport and metabolism	G	COG0697
PANA_1559	METABOLISM	Amino acid transport and metabolism	E	COG0697
PANA_1559	POORLY CHARACTERIZED	General function prediction only	R	COG0697
PANA_1560	METABOLISM	Nucleotide transport and metabolism	F	COG0518
PANA_1561	METABOLISM	Lipid transport and metabolism	I	COG3963
PANA_1562	METABOLISM	Energy production and conversion	C	COG1012
PANA_1573	METABOLISM	Energy production and conversion	C	COG1902
PANA_1575	METABOLISM	Energy production and conversion	C	COG3053
PANA_1582	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG4977
PANA_1610	NO COG	No COG association	NC	-
PANA_1613	NO COG	No COG association	NC	-
PANA_1614	NO COG	No COG association	NC	-
PANA_1615	METABOLISM	Amino acid transport and metabolism	E	COG0620
PANA_1617	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0386
PANA_1618	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1846
PANA_1619	METABOLISM	Coenzyme transport and metabolism	H	COG0163
PANA_1620	METABOLISM	Coenzyme transport and metabolism	H	COG0043
PANA_1622	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_1625	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_1634	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0012
PANA_1635	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0193
PANA_1636	NO COG	No COG association	NC	-
PANA_1637	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2199
PANA_1638	METABOLISM	Nucleotide transport and metabolism	F	COG0462
PANA_1638	METABOLISM	Amino acid transport and metabolism	E	COG0462
PANA_1639	METABOLISM	Lipid transport and metabolism	I	COG1947
PANA_1640	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3017
PANA_1641	METABOLISM	Coenzyme transport and metabolism	H	COG0373
PANA_1642	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0216
PANA_1643	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2890
PANA_1644	POORLY CHARACTERIZED	Function unknown	S	COG3094
PANA_1645	POORLY CHARACTERIZED	Function unknown	S	COG2912
PANA_1646	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2877
PANA_1647	METABOLISM	Inorganic ion transport and metabolism	P	COG0387
PANA_1648	POORLY CHARACTERIZED	General function prediction only	R	COG4572
PANA_1649	METABOLISM	Inorganic ion transport and metabolism	P	COG3703
PANA_1650	POORLY CHARACTERIZED	General function prediction only	R	COG0515
PANA_1650	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0515
PANA_1650	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0515
PANA_1650	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0515
PANA_1651	NO COG	No COG association	NC	-
PANA_1652	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0631
PANA_1653	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3456
PANA_1654	POORLY CHARACTERIZED	Function unknown	S	COG3913
PANA_1655	POORLY CHARACTERIZED	Function unknown	S	COG3523
PANA_1656	POORLY CHARACTERIZED	Function unknown	S	COG3455
PANA_1661	METABOLISM	Inorganic ion transport and metabolism	P	COG1116
PANA_1662	METABOLISM	Energy production and conversion	C	COG1251
PANA_1665	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1349
PANA_1665	METABOLISM	Carbohydrate transport and metabolism	G	COG1349
PANA_1666	METABOLISM	Lipid transport and metabolism	I	COG1028
PANA_1666	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1028
PANA_1666	POORLY CHARACTERIZED	General function prediction only	R	COG1028
PANA_1667	METABOLISM	Carbohydrate transport and metabolism	G	COG1879
PANA_1668	POORLY CHARACTERIZED	General function prediction only	R	COG5618
PANA_1669	METABOLISM	Carbohydrate transport and metabolism	G	COG1129
PANA_1670	METABOLISM	Carbohydrate transport and metabolism	G	COG1172
PANA_1671	METABOLISM	Lipid transport and metabolism	I	COG1028
PANA_1671	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1028
PANA_1671	POORLY CHARACTERIZED	General function prediction only	R	COG1028
PANA_1672	NO COG	No COG association	NC	-
PANA_1673	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3203
PANA_1674	METABOLISM	Amino acid transport and metabolism	E	COG4992
PANA_1675	METABOLISM	Amino acid transport and metabolism	E	COG3138

PANA_1676	METABOLISM	Energy production and conversion	C	COG1012
PANA_1677	METABOLISM	Amino acid transport and metabolism	E	COG3724
PANA_1678	METABOLISM	Amino acid transport and metabolism	E	COG2988
PANA_1679	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG3678
PANA_1679	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG3678
PANA_1679	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3678
PANA_1679	METABOLISM	Inorganic ion transport and metabolism	P	COG3678
PANA_1680	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0322
PANA_1681	NO COG	No COG association	NC	-
PANA_1683	METABOLISM	Coenzyme transport and metabolism	H	COG0171
PANA_1684	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_1685	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2913
PANA_1687	POORLY CHARACTERIZED	General function prediction only	R	COG1823
PANA_1688	POORLY CHARACTERIZED	General function prediction only	R	COG1988
PANA_1689	POORLY CHARACTERIZED	General function prediction only	R	COG0637
PANA_1690	NO COG	No COG association	NC	-
PANA_1691	METABOLISM	Carbohydrate transport and metabolism	G	COG3001
PANA_1692	NO COG	No COG association	NC	-
PANA_1693	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3137
PANA_1694	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0441
PANA_1695	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0290
PANA_1697	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0291
PANA_1696	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0292
PANA_1698	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0016
PANA_1699	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0072
PANA_1700	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0776
PANA_1702	METABOLISM	Coenzyme transport and metabolism	H	COG4139
PANA_1703	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0386
PANA_1704	METABOLISM	Coenzyme transport and metabolism	H	COG4138
PANA_1705	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0791
PANA_1706	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2200
PANA_1707	POORLY CHARACTERIZED	Function unknown	S	COG0397
PANA_1708	METABOLISM	Amino acid transport and metabolism	E	COG0722
PANA_1709	POORLY CHARACTERIZED	Function unknown	S	COG1806
PANA_1710	METABOLISM	Carbohydrate transport and metabolism	G	COG0574
PANA_1711	POORLY CHARACTERIZED	General function prediction only	R	COG0628
PANA_1712	METABOLISM	Energy production and conversion	C	COG0277
PANA_1714	POORLY CHARACTERIZED	Function unknown	S	COG0316
PANA_1715	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0719
PANA_1716	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0396
PANA_1717	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0719
PANA_1718	METABOLISM	Amino acid transport and metabolism	E	COG0520
PANA_1719	POORLY CHARACTERIZED	General function prediction only	R	COG2166
PANA_1720	POORLY CHARACTERIZED	Function unknown	S	COG1376
PANA_1721	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG4238
PANA_1722	METABOLISM	Carbohydrate transport and metabolism	G	COG0469
PANA_1723	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0534
PANA_1724	METABOLISM	Coenzyme transport and metabolism	H	COG0307
PANA_1726	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2230
PANA_1727	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_1728	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_1729	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_1731	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0791
PANA_1732	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0278
PANA_1733	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0847
PANA_1734	METABOLISM	Amino acid transport and metabolism	E	COG0346
PANA_1735	METABOLISM	Energy production and conversion	C	COG1902
PANA_1736	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1309
PANA_1738	POORLY CHARACTERIZED	General function prediction only	R	COG4989
PANA_1739	METABOLISM	Inorganic ion transport and metabolism	P	COG2032
PANA_1740	NO COG	No COG association	NC	-
PANA_1741	POORLY CHARACTERIZED	Function unknown	S	COG1289
PANA_1742	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1566
PANA_1743	NO COG	No COG association	NC	-

PANA_1744	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1846
PANA_1745	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3133
PANA_1746	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG2377
PANA_1748	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2200
PANA_1750	METABOLISM	Carbohydrate transport and metabolism	G	COG1523
PANA_1751	METABOLISM	Carbohydrate transport and metabolism	G	COG3280
PANA_1752	METABOLISM	Carbohydrate transport and metabolism	G	COG0296
PANA_1753	METABOLISM	Inorganic ion transport and metabolism	P	COG0753
PANA_1757	NO COG	No COG association	NC	-
PANA_1763	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1974
PANA_1763	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1974
PANA_1766	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1309
PANA_1768	NO COG	No COG association	NC	-
PANA_1770	NO COG	No COG association	NC	-
PANA_1771	NO COG	No COG association	NC	-
PANA_1773	POORLY CHARACTERIZED	General function prediction only	R	COG3895
PANA_1774	METABOLISM	Coenzyme transport and metabolism	H	COG0259
PANA_1775	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0162
PANA_1776	METABOLISM	Coenzyme transport and metabolism	H	COG2240
PANA_1777	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0625
PANA_1778	METABOLISM	Amino acid transport and metabolism	E	COG3104
PANA_1779	METABOLISM	Nucleotide transport and metabolism	F	COG0504
PANA_1780	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0177
PANA_1781	METABOLISM	Energy production and conversion	C	COG4660
PANA_1782	METABOLISM	Energy production and conversion	C	COG4659
PANA_1783	METABOLISM	Energy production and conversion	C	COG4658
PANA_1784	METABOLISM	Energy production and conversion	C	COG4656
PANA_1785	METABOLISM	Energy production and conversion	C	COG2878
PANA_1786	METABOLISM	Energy production and conversion	C	COG4657
PANA_1787	NO COG	No COG association	NC	-
PANA_1788	NO COG	No COG association	NC	-
PANA_1789	METABOLISM	Carbohydrate transport and metabolism	G	COG1879
PANA_1790	METABOLISM	Carbohydrate transport and metabolism	G	COG1129
PANA_1791	METABOLISM	Carbohydrate transport and metabolism	G	COG1172
PANA_1792	METABOLISM	Carbohydrate transport and metabolism	G	COG1172
PANA_1815	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1278
PANA_1816	POORLY CHARACTERIZED	Function unknown	S	COG3122
PANA_1819	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0564
PANA_1823	METABOLISM	Carbohydrate transport and metabolism	G	COG2160
PANA_1824	METABOLISM	Energy production and conversion	C	COG1069
PANA_1825	METABOLISM	Carbohydrate transport and metabolism	G	COG1879
PANA_1826	METABOLISM	Carbohydrate transport and metabolism	G	COG1129
PANA_1827	METABOLISM	Carbohydrate transport and metabolism	G	COG1172
PANA_1828	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_1829	POORLY CHARACTERIZED	General function prediction only	R	COG0673
PANA_1830	METABOLISM	Nucleotide transport and metabolism	F	COG1816
PANA_1831	POORLY CHARACTERIZED	Function unknown	S	COG5339
PANA_1832	METABOLISM	Carbohydrate transport and metabolism	G	COG1482
PANA_1833	METABOLISM	Energy production and conversion	C	COG0114
PANA_1834	NO COG	No COG association	NC	-
PANA_1835	NO COG	No COG association	NC	-
PANA_1837	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0841
PANA_1838	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0845
PANA_1839	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_1842	POORLY CHARACTERIZED	Function unknown	S	COG4458
PANA_1843	POORLY CHARACTERIZED	Function unknown	S	COG4457
PANA_1844	NO COG	No COG association	NC	-
PANA_1845	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG3135
PANA_1846	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1396
PANA_1847	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1335
PANA_1849	METABOLISM	Inorganic ion transport and metabolism	P	COG1629
PANA_1850	METABOLISM	Lipid transport and metabolism	I	COG2084
PANA_1851	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_1852	METABOLISM	Amino acid transport and metabolism	E	COG0119

PANA_1853	METABOLISM	Amino acid transport and metabolism	E	COG2355
PANA_1854	POORLY CHARACTERIZED	General function prediction only	R	COG1235
PANA_1855	METABOLISM	Coenzyme transport and metabolism	H	COG5424
PANA_1857	POORLY CHARACTERIZED	General function prediction only	R	COG0535
PANA_1858	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1025
PANA_1859	POORLY CHARACTERIZED	Function unknown	S	COG2353
PANA_1860	METABOLISM	Carbohydrate transport and metabolism	G	COG0205
PANA_1862	POORLY CHARACTERIZED	Function unknown	S	COG3037
PANA_1863	METABOLISM	Carbohydrate transport and metabolism	G	COG0191
PANA_1865	METABOLISM	Carbohydrate transport and metabolism	G	COG1762
PANA_1865	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1762
PANA_1866	METABOLISM	Nucleotide transport and metabolism	F	COG0563
PANA_1870	POORLY CHARACTERIZED	General function prediction only	R	COG0110
PANA_1891	POORLY CHARACTERIZED	Function unknown	S	COG3314
PANA_1892	POORLY CHARACTERIZED	Function unknown	S	COG0700
PANA_1893	POORLY CHARACTERIZED	General function prediction only	R	COG1473
PANA_1901	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1670
PANA_1902	NO COG	No COG association	NC	-
PANA_1903	NO COG	No COG association	NC	-
PANA_1904	METABOLISM	Energy production and conversion	C	COG1902
PANA_1909	METABOLISM	Carbohydrate transport and metabolism	G	COG3833
PANA_1910	METABOLISM	Carbohydrate transport and metabolism	G	COG1175
PANA_1911	METABOLISM	Carbohydrate transport and metabolism	G	COG2182
PANA_1912	METABOLISM	Carbohydrate transport and metabolism	G	COG3839
PANA_1913	METABOLISM	Carbohydrate transport and metabolism	G	COG4580
PANA_1914	NO COG	No COG association	NC	-
PANA_1917	METABOLISM	Carbohydrate transport and metabolism	G	COG1640
PANA_1918	METABOLISM	Carbohydrate transport and metabolism	G	COG0058
PANA_1919	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2909
PANA_1922	POORLY CHARACTERIZED	Function unknown	S	COG1649
PANA_1951	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1476
PANA_1957	METABOLISM	Lipid transport and metabolism	I	COG1028
PANA_1957	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1028
PANA_1957	POORLY CHARACTERIZED	General function prediction only	R	COG1028
PANA_1958	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1680
PANA_1959	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0438
PANA_1960	METABOLISM	Coenzyme transport and metabolism	H	COG0654
PANA_1960	METABOLISM	Energy production and conversion	C	COG0654
PANA_1961	METABOLISM	Amino acid transport and metabolism	E	COG0339
PANA_1962	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1975
PANA_1963	METABOLISM	Nucleotide transport and metabolism	F	COG4631
PANA_1964	METABOLISM	Nucleotide transport and metabolism	F	COG0402
PANA_1964	POORLY CHARACTERIZED	General function prediction only	R	COG0402
PANA_1965	POORLY CHARACTERIZED	General function prediction only	R	COG4221
PANA_1966	POORLY CHARACTERIZED	Function unknown	S	COG1742
PANA_1967	NO COG	No COG association	NC	-
PANA_1968	NO COG	No COG association	NC	-
PANA_1969	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_1970	METABOLISM	Amino acid transport and metabolism	E	COG1174
PANA_1971	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1732
PANA_1972	METABOLISM	Amino acid transport and metabolism	E	COG1174
PANA_1973	METABOLISM	Amino acid transport and metabolism	E	COG1125
PANA_1974	METABOLISM	Coenzyme transport and metabolism	H	COG0132
PANA_1975	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1940
PANA_1975	METABOLISM	Carbohydrate transport and metabolism	G	COG1940
PANA_1976	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_1977	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_1979	METABOLISM	Energy production and conversion	C	COG3038
PANA_1980	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_1982	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1734
PANA_1983	METABOLISM	Amino acid transport and metabolism	E	COG0747
PANA_1984	METABOLISM	Amino acid transport and metabolism	E	COG2195
PANA_1985	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_1985	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840

PANA_1986	METABOLISM	Lipid transport and metabolism	I	COG2272
PANA_1988	POORLY CHARACTERIZED	Function unknown	S	COG5383
PANA_1989	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_1990	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_1991	POORLY CHARACTERIZED	Function unknown	S	COG2840
PANA_1992	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0664
PANA_1993	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0589
PANA_1994	METABOLISM	Energy production and conversion	C	COG1282
PANA_1995	METABOLISM	Energy production and conversion	C	COG3288
PANA_1996	NO COG	No COG association	NC	-
PANA_1997	METABOLISM	Amino acid transport and metabolism	E	COG0531
PANA_1999	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_1999	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_2000	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_2001	METABOLISM	Amino acid transport and metabolism	E	COG2317
PANA_2002	NO COG	No COG association	NC	-
PANA_2005	METABOLISM	Amino acid transport and metabolism	E	COG3591
PANA_2006	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1643
PANA_2007	METABOLISM	Lipid transport and metabolism	I	COG1182
PANA_2008	POORLY CHARACTERIZED	Function unknown	S	COG3784
PANA_2010	NO COG	No COG association	NC	-
PANA_2011	METABOLISM	Energy production and conversion	C	COG1052
PANA_2011	METABOLISM	Coenzyme transport and metabolism	H	COG1052
PANA_2011	POORLY CHARACTERIZED	General function prediction only	R	COG1052
PANA_2012	POORLY CHARACTERIZED	General function prediction only	R	COG3042
PANA_2014	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0037
PANA_2015	METABOLISM	Inorganic ion transport and metabolism	P	COG0598
PANA_2016	METABOLISM	Amino acid transport and metabolism	E	COG4166
PANA_2017	METABOLISM	Amino acid transport and metabolism	E	COG2866
PANA_2018	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG4948
PANA_2018	POORLY CHARACTERIZED	General function prediction only	R	COG4948
PANA_2019	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG2077
PANA_2021	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3283
PANA_2021	METABOLISM	Amino acid transport and metabolism	E	COG3283
PANA_2022	POORLY CHARACTERIZED	Function unknown	S	COG3768
PANA_2023	POORLY CHARACTERIZED	General function prediction only	R	COG3106
PANA_2024	NO COG	No COG association	NC	-
PANA_2025	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1983
PANA_2025	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1983
PANA_2026	NO COG	No COG association	NC	-
PANA_2027	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1842
PANA_2027	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1842
PANA_2028	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1221
PANA_2028	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1221
PANA_2030	METABOLISM	Amino acid transport and metabolism	E	COG4166
PANA_2031	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG4168
PANA_2032	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG4171
PANA_2033	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG4170
PANA_2034	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG4167
PANA_2035	METABOLISM	Lipid transport and metabolism	I	COG0623
PANA_2036	POORLY CHARACTERIZED	Function unknown	S	COG4950
PANA_2037	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_2038	METABOLISM	Carbohydrate transport and metabolism	G	COG1263
PANA_2039	METABOLISM	Energy production and conversion	C	COG0243
PANA_2040	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG4776
PANA_2041	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1966
PANA_2042	POORLY CHARACTERIZED	Function unknown	S	COG2879
PANA_2043	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG5001
PANA_2044	METABOLISM	Lipid transport and metabolism	I	COG1024
PANA_2045	METABOLISM	Carbohydrate transport and metabolism	G	COG0235
PANA_2046	NO COG	No COG association	NC	-
PANA_2047	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0023
PANA_2048	METABOLISM	Nucleotide transport and metabolism	F	COG0284
PANA_2049	METABOLISM	Carbohydrate transport and metabolism	G	COG2956

PANA_2050	POORLY CHARACTERIZED	Function unknown	S	COG3771
PANA_2051	METABOLISM	Lipid transport and metabolism	I	COG0671
PANA_2052	METABOLISM	Coenzyme transport and metabolism	H	COG0807
PANA_2053	METABOLISM	Energy production and conversion	C	COG1048
PANA_2055	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_2056	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0550
PANA_2057	NO COG	No COG association	NC	-
PANA_2058	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0616
PANA_2058	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0616
PANA_2059	METABOLISM	Lipid transport and metabolism	I	COG1028
PANA_2059	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1028
PANA_2059	POORLY CHARACTERIZED	General function prediction only	R	COG1028
PANA_2060	METABOLISM	Coenzyme transport and metabolism	H	COG2109
PANA_2061	METABOLISM	Energy production and conversion	C	COG2851
PANA_2062	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1187
PANA_2063	NO COG	No COG association	NC	-
PANA_2064	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0009
PANA_2065	POORLY CHARACTERIZED	General function prediction only	R	COG0613
PANA_2066	METABOLISM	Amino acid transport and metabolism	E	COG0147
PANA_2066	METABOLISM	Coenzyme transport and metabolism	H	COG0147
PANA_2067	METABOLISM	Amino acid transport and metabolism	E	COG0547
PANA_2068	METABOLISM	Amino acid transport and metabolism	E	COG0134
PANA_2069	METABOLISM	Amino acid transport and metabolism	E	COG0133
PANA_2070	METABOLISM	Amino acid transport and metabolism	E	COG0159
PANA_2071	POORLY CHARACTERIZED	General function prediction only	R	COG2823
PANA_2072	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3047
PANA_2073	POORLY CHARACTERIZED	General function prediction only	R	COG0727
PANA_2074	NO COG	No COG association	NC	-
PANA_2075	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG2917
PANA_2076	METABOLISM	Lipid transport and metabolism	I	COG1607
PANA_2077	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0810
PANA_2078	POORLY CHARACTERIZED	Function unknown	S	COG2350
PANA_2081	METABOLISM	Lipid transport and metabolism	I	COG1502
PANA_2082	POORLY CHARACTERIZED	Function unknown	S	COG3099
PANA_2083	METABOLISM	Amino acid transport and metabolism	E	COG4608
PANA_2084	METABOLISM	Amino acid transport and metabolism	E	COG0444
PANA_2084	METABOLISM	Inorganic ion transport and metabolism	P	COG0444
PANA_2085	METABOLISM	Amino acid transport and metabolism	E	COG1173
PANA_2085	METABOLISM	Inorganic ion transport and metabolism	P	COG1173
PANA_2086	METABOLISM	Amino acid transport and metabolism	E	COG0601
PANA_2086	METABOLISM	Inorganic ion transport and metabolism	P	COG0601
PANA_2087	METABOLISM	Amino acid transport and metabolism	E	COG4166
PANA_2088	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG2095
PANA_2089	METABOLISM	Energy production and conversion	C	COG1454
PANA_2090	METABOLISM	Nucleotide transport and metabolism	F	COG1435
PANA_2091	POORLY CHARACTERIZED	General function prediction only	R	COG2916
PANA_2092	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0451
PANA_2092	METABOLISM	Carbohydrate transport and metabolism	G	COG0451
PANA_2093	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1004
PANA_2094	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1210
PANA_2095	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2204
PANA_2096	POORLY CHARACTERIZED	General function prediction only	R	COG1752
PANA_2098	METABOLISM	Nucleotide transport and metabolism	F	COG0788
PANA_2100	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0708
PANA_2101	NO COG	No COG association	NC	-
PANA_2102	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0550
PANA_2103	METABOLISM	Energy production and conversion	C	COG0778
PANA_2104	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0616
PANA_2104	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0616
PANA_2105	METABOLISM	Amino acid transport and metabolism	E	COG0252
PANA_2105	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0252
PANA_2106	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1335
PANA_2107	POORLY CHARACTERIZED	Function unknown	S	COG3139
PANA_2109	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0229

PANA_2110	METABOLISM	Carbohydrate transport and metabolism	G	COG0057
PANA_2111	METABOLISM	Carbohydrate transport and metabolism	G	COG0676
PANA_2112	POORLY CHARACTERIZED	General function prediction only	R	COG0656
PANA_2113	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3713
PANA_2114	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2766
PANA_2115	POORLY CHARACTERIZED	Function unknown	S	COG2718
PANA_2116	METABOLISM	Inorganic ion transport and metabolism	P	COG3263
PANA_2117	NO COG	No COG association	NC	-
PANA_2118	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2148
PANA_2119	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_2120	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0787
PANA_2121	METABOLISM	Amino acid transport and metabolism	E	COG0665
PANA_2122	POORLY CHARACTERIZED	Function unknown	S	COG2719
PANA_2123	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2186
PANA_2124	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1495
PANA_2125	POORLY CHARACTERIZED	Function unknown	S	COG2983
PANA_2127	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG0179
PANA_2128	POORLY CHARACTERIZED	Function unknown	S	COG3100
PANA_2129	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0850
PANA_2130	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG2894
PANA_2131	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0851
PANA_2132	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0349
PANA_2133	METABOLISM	Lipid transport and metabolism	I	COG0318
PANA_2133	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG0318
PANA_2134	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3065
PANA_2135	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1214
PANA_2136	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1199
PANA_2136	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1199
PANA_2138	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0251
PANA_2139	POORLY CHARACTERIZED	Function unknown	S	COG3140
PANA_2140	METABOLISM	Amino acid transport and metabolism	E	COG0147
PANA_2140	METABOLISM	Coenzyme transport and metabolism	H	COG0147
PANA_2141	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0494
PANA_2141	POORLY CHARACTERIZED	General function prediction only	R	COG0494
PANA_2142	METABOLISM	Amino acid transport and metabolism	E	COG1760
PANA_2143	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2200
PANA_2144	METABOLISM	Inorganic ion transport and metabolism	P	COG0861
PANA_2145	METABOLISM	Carbohydrate transport and metabolism	G	COG3444
PANA_2146	METABOLISM	Carbohydrate transport and metabolism	G	COG3715
PANA_2147	METABOLISM	Carbohydrate transport and metabolism	G	COG3716
PANA_2148	POORLY CHARACTERIZED	Function unknown	S	COG4811
PANA_2149	METABOLISM	Coenzyme transport and metabolism	H	COG2226
PANA_2152	NO COG	No COG association	NC	-
PANA_2154	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1414
PANA_2155	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_2156	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0501
PANA_2157	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0793
PANA_2158	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3109
PANA_2159	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1956
PANA_2160	POORLY CHARACTERIZED	Function unknown	S	COG2995
PANA_2161	POORLY CHARACTERIZED	General function prediction only	R	COG3008
PANA_2162	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0144
PANA_2164	NO COG	No COG association	NC	-
PANA_2165	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1292
PANA_2166	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1309
PANA_2167	METABOLISM	Energy production and conversion	C	COG1012
PANA_2168	METABOLISM	Amino acid transport and metabolism	E	COG2303
PANA_2182	NO COG	No COG association	NC	-
PANA_2183	METABOLISM	Inorganic ion transport and metabolism	P	COG1276
PANA_2185	NO COG	No COG association	NC	-
PANA_2186	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0847
PANA_2187	METABOLISM	Amino acid transport and metabolism	E	COG1770
PANA_2188	POORLY CHARACTERIZED	Function unknown	S	COG2979
PANA_2189	METABOLISM	Nucleotide transport and metabolism	F	COG0027

PANA_2190	METABOLISM	Carbohydrate transport and metabolism	G	COG0800
PANA_2191	METABOLISM	Carbohydrate transport and metabolism	G	COG0364
PANA_2192	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1737
PANA_2193	METABOLISM	Carbohydrate transport and metabolism	G	COG0469
PANA_2194	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_2195	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1560
PANA_2196	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0739
PANA_2197	METABOLISM	Inorganic ion transport and metabolism	P	COG4531
PANA_2198	METABOLISM	Inorganic ion transport and metabolism	P	COG1121
PANA_2199	METABOLISM	Inorganic ion transport and metabolism	P	COG1108
PANA_2200	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2255
PANA_2201	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0632
PANA_2202	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0817
PANA_2203	POORLY CHARACTERIZED	Function unknown	S	COG0217
PANA_2205	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0494
PANA_2205	POORLY CHARACTERIZED	General function prediction only	R	COG0494
PANA_2206	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0173
PANA_2208	POORLY CHARACTERIZED	Function unknown	S	COG1801
PANA_2210	POORLY CHARACTERIZED	General function prediction only	R	COG4106
PANA_2211	METABOLISM	Coenzyme transport and metabolism	H	COG2227
PANA_2212	METABOLISM	Amino acid transport and metabolism	E	COG1280
PANA_2213	METABOLISM	Energy production and conversion	C	COG0371
PANA_2214	NO COG	No COG association	NC	-
PANA_2215	METABOLISM	Inorganic ion transport and metabolism	P	COG3142
PANA_2216	METABOLISM	Amino acid transport and metabolism	E	COG1168
PANA_2217	POORLY CHARACTERIZED	Function unknown	S	COG3102
PANA_2218	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0018
PANA_2219	METABOLISM	Carbohydrate transport and metabolism	G	COG0580
PANA_2220	NO COG	No COG association	NC	-
PANA_2221	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1298
PANA_2221	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1298
PANA_2222	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1377
PANA_2222	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1377
PANA_2223	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG3143
PANA_2223	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3143
PANA_2224	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2204
PANA_2225	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG2201
PANA_2225	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2201
PANA_2226	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1352
PANA_2226	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1352
PANA_2227	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_2227	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840
PANA_2228	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_2228	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840
PANA_2229	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0835
PANA_2229	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0835
PANA_2230	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0643
PANA_2230	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0643
PANA_2231	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1360
PANA_2232	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1291
PANA_2233	NO COG	No COG association	NC	-
PANA_2234	NO COG	No COG association	NC	-
PANA_2235	METABOLISM	Carbohydrate transport and metabolism	G	COG0380
PANA_2236	METABOLISM	Carbohydrate transport and metabolism	G	COG1877
PANA_2237	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_2237	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_2238	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_2239	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0225
PANA_2243	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2204
PANA_2244	METABOLISM	Energy production and conversion	C	COG2141
PANA_2245	POORLY CHARACTERIZED	Function unknown	S	COG4529
PANA_2246	METABOLISM	Energy production and conversion	C	COG2141
PANA_2247	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_2248	METABOLISM	Lipid transport and metabolism	I	COG1960

PANA_2249	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1702
PANA_2253	METABOLISM	Amino acid transport and metabolism	E	COG0591
PANA_2253	POORLY CHARACTERIZED	General function prediction only	R	COG0591
PANA_2254	METABOLISM	Energy production and conversion	C	COG4230
PANA_2256	POORLY CHARACTERIZED	Function unknown	S	COG4763
PANA_2259	METABOLISM	Amino acid transport and metabolism	E	COG1126
PANA_2260	METABOLISM	Amino acid transport and metabolism	E	COG0765
PANA_2261	METABOLISM	Amino acid transport and metabolism	E	COG0834
PANA_2261	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0834
PANA_2262	METABOLISM	Amino acid transport and metabolism	E	COG2515
PANA_2263	NO COG	No COG association	NC	-
PANA_2264	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1191
PANA_2265	NO COG	No COG association	NC	-
PANA_2266	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1209
PANA_2282	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1345
PANA_2283	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1516
PANA_2283	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1516
PANA_2283	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1516
PANA_2284	NO COG	No COG association	NC	-
PANA_2285	METABOLISM	Carbohydrate transport and metabolism	G	COG0366
PANA_2286	NO COG	No COG association	NC	-
PANA_2289	NO COG	No COG association	NC	-
PANA_2290	NO COG	No COG association	NC	-
PANA_2294	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1677
PANA_2294	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1677
PANA_2295	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1766
PANA_2295	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1766
PANA_2296	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1536
PANA_2297	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1317
PANA_2297	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1317
PANA_2298	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1157
PANA_2298	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1157
PANA_2299	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG2882
PANA_2299	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG2882
PANA_2299	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG2882
PANA_2300	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG3144
PANA_2301	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1580
PANA_2302	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1868
PANA_2303	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1886
PANA_2303	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1886
PANA_2304	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG3190
PANA_2305	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1338
PANA_2305	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1338
PANA_2306	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1987
PANA_2306	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1987
PANA_2307	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG1684
PANA_2307	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1684
PANA_2308	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2197
PANA_2308	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2197
PANA_2309	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_2309	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840
PANA_2310	NO COG	No COG association	NC	-
PANA_2313	POORLY CHARACTERIZED	General function prediction only	R	COG3769
PANA_2314	NO COG	No COG association	NC	-
PANA_2317	POORLY CHARACTERIZED	Function unknown	S	COG2354
PANA_2319	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG3727
PANA_2321	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0270
PANA_2324	POORLY CHARACTERIZED	General function prediction only	R	COG1418
PANA_2325	NO COG	No COG association	NC	-
PANA_2326	POORLY CHARACTERIZED	Function unknown	S	COG3228
PANA_2331	METABOLISM	Coenzyme transport and metabolism	H	COG0276
PANA_2333	POORLY CHARACTERIZED	General function prediction only	R	COG2270
PANA_2339	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0071
PANA_2352	POORLY CHARACTERIZED	Function unknown	S	COG3501

PANA_2353	POORLY CHARACTERIZED	General function prediction only	R	COG0515
PANA_2353	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0515
PANA_2353	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0515
PANA_2353	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0515
PANA_2354	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0542
PANA_2355	POORLY CHARACTERIZED	Function unknown	S	COG3520
PANA_2356	POORLY CHARACTERIZED	Function unknown	S	COG3519
PANA_2357	POORLY CHARACTERIZED	Function unknown	S	COG3518
PANA_2358	POORLY CHARACTERIZED	General function prediction only	R	COG4455
PANA_2359	NO COG	No COG association	NC	-
PANA_2360	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0631
PANA_2361	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3456
PANA_2364	POORLY CHARACTERIZED	Function unknown	S	COG3157
PANA_2365	POORLY CHARACTERIZED	Function unknown	S	COG3517
PANA_2366	POORLY CHARACTERIZED	Function unknown	S	COG3516
PANA_2367	POORLY CHARACTERIZED	Function unknown	S	COG3515
PANA_2368	POORLY CHARACTERIZED	Function unknown	S	COG3913
PANA_2369	POORLY CHARACTERIZED	Function unknown	S	COG3523
PANA_2370	POORLY CHARACTERIZED	Function unknown	S	COG3455
PANA_2371	POORLY CHARACTERIZED	Function unknown	S	COG3522
PANA_2372	POORLY CHARACTERIZED	Function unknown	S	COG3521
PANA_2373	NO COG	No COG association	NC	-
PANA_2374	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0534
PANA_2375	METABOLISM	Nucleotide transport and metabolism	F	COG0775
PANA_2377	POORLY CHARACTERIZED	General function prediction only	R	COG2252
PANA_2378	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_2378	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_2381	POORLY CHARACTERIZED	General function prediction only	R	COG3822
PANA_2382	METABOLISM	Carbohydrate transport and metabolism	G	COG1879
PANA_2383	POORLY CHARACTERIZED	General function prediction only	R	COG4158
PANA_2384	METABOLISM	Carbohydrate transport and metabolism	G	COG1129
PANA_2385	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG5000
PANA_2386	METABOLISM	Inorganic ion transport and metabolism	P	COG3639
PANA_2387	METABOLISM	Inorganic ion transport and metabolism	P	COG3639
PANA_2388	METABOLISM	Inorganic ion transport and metabolism	P	COG3221
PANA_2389	METABOLISM	Inorganic ion transport and metabolism	P	COG3638
PANA_2391	METABOLISM	Inorganic ion transport and metabolism	P	COG3709
PANA_2392	METABOLISM	Inorganic ion transport and metabolism	P	COG3454
PANA_2393	METABOLISM	Inorganic ion transport and metabolism	P	COG4778
PANA_2394	METABOLISM	Inorganic ion transport and metabolism	P	COG4107
PANA_2395	METABOLISM	Inorganic ion transport and metabolism	P	COG3627
PANA_2396	METABOLISM	Inorganic ion transport and metabolism	P	COG3626
PANA_2397	METABOLISM	Inorganic ion transport and metabolism	P	COG3625
PANA_2398	METABOLISM	Inorganic ion transport and metabolism	P	COG3624
PANA_2399	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2188
PANA_2403	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1522
PANA_2404	METABOLISM	Amino acid transport and metabolism	E	COG0160
PANA_2405	POORLY CHARACTERIZED	General function prediction only	R	COG1473
PANA_2406	METABOLISM	Coenzyme transport and metabolism	H	COG0111
PANA_2406	METABOLISM	Amino acid transport and metabolism	E	COG0111
PANA_2407	POORLY CHARACTERIZED	General function prediction only	R	COG3153
PANA_2408	METABOLISM	Amino acid transport and metabolism	E	COG0665
PANA_2409	POORLY CHARACTERIZED	General function prediction only	R	COG1123
PANA_2410	METABOLISM	Amino acid transport and metabolism	E	COG1173
PANA_2410	METABOLISM	Inorganic ion transport and metabolism	P	COG1173
PANA_2411	METABOLISM	Amino acid transport and metabolism	E	COG0601
PANA_2411	METABOLISM	Inorganic ion transport and metabolism	P	COG0601
PANA_2412	METABOLISM	Amino acid transport and metabolism	E	COG0747
PANA_2413	METABOLISM	Amino acid transport and metabolism	E	COG0665
PANA_2414	POORLY CHARACTERIZED	General function prediction only	R	COG1011
PANA_2415	METABOLISM	Energy production and conversion	C	COG1012
PANA_2416	METABOLISM	Energy production and conversion	C	COG1012
PANA_2417	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_2418	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583

PANA_2419	METABOLISM	Carbohydrate transport and metabolism	G	COG1172
PANA_2420	METABOLISM	Carbohydrate transport and metabolism	G	COG1129
PANA_2421	METABOLISM	Carbohydrate transport and metabolism	G	COG1879
PANA_2427	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1522
PANA_2429	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1309
PANA_2430	METABOLISM	Energy production and conversion	C	COG0277
PANA_2431	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1442
PANA_2432	POORLY CHARACTERIZED	General function prediction only	R	COG0673
PANA_2435	POORLY CHARACTERIZED	Function unknown	S	COG3254
PANA_2436	METABOLISM	Carbohydrate transport and metabolism	G	COG0235
PANA_2437	METABOLISM	Carbohydrate transport and metabolism	G	COG4806
PANA_2438	METABOLISM	Carbohydrate transport and metabolism	G	COG1070
PANA_2439	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG4977
PANA_2440	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_2441	NO COG	No COG association	NC	-
PANA_2443	METABOLISM	Energy production and conversion	C	COG1012
PANA_2448	METABOLISM	Energy production and conversion	C	COG2141
PANA_2449	METABOLISM	Coenzyme transport and metabolism	H	COG1893
PANA_2450	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_2451	POORLY CHARACTERIZED	Function unknown	S	COG2926
PANA_2452	POORLY CHARACTERIZED	Function unknown	S	COG1289
PANA_2453	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1686
PANA_2454	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2925
PANA_2456	METABOLISM	Carbohydrate transport and metabolism	G	COG3622
PANA_2457	METABOLISM	Carbohydrate transport and metabolism	G	COG0235
PANA_2458	POORLY CHARACTERIZED	Function unknown	S	COG3395
PANA_2460	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1349
PANA_2460	METABOLISM	Carbohydrate transport and metabolism	G	COG1349
PANA_2461	METABOLISM	Amino acid transport and metabolism	E	COG0531
PANA_2462	METABOLISM	Amino acid transport and metabolism	E	COG0174
PANA_2463	POORLY CHARACTERIZED	General function prediction only	R	COG2071
PANA_2464	METABOLISM	Carbohydrate transport and metabolism	G	COG0662
PANA_2465	METABOLISM	Amino acid transport and metabolism	E	COG0665
PANA_2466	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0451
PANA_2466	METABOLISM	Carbohydrate transport and metabolism	G	COG0451
PANA_2467	METABOLISM	Amino acid transport and metabolism	E	COG0040
PANA_2468	METABOLISM	Amino acid transport and metabolism	E	COG0141
PANA_2469	METABOLISM	Amino acid transport and metabolism	E	COG0079
PANA_2471	METABOLISM	Amino acid transport and metabolism	E	COG0131
PANA_2472	METABOLISM	Amino acid transport and metabolism	E	COG0118
PANA_2473	METABOLISM	Amino acid transport and metabolism	E	COG0106
PANA_2474	METABOLISM	Amino acid transport and metabolism	E	COG0107
PANA_2475	METABOLISM	Amino acid transport and metabolism	E	COG0139
PANA_2478	POORLY CHARACTERIZED	Function unknown	S	COG2606
PANA_2479	METABOLISM	Amino acid transport and metabolism	E	COG0834
PANA_2479	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0834
PANA_2480	POORLY CHARACTERIZED	General function prediction only	R	COG3608
PANA_2481	METABOLISM	Carbohydrate transport and metabolism	G	COG0362
PANA_2493	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1087
PANA_2494	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1210
PANA_2495	POORLY CHARACTERIZED	General function prediction only	R	COG2244
PANA_2496	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0438
PANA_2497	POORLY CHARACTERIZED	Function unknown	S	COG2327
PANA_2498	NO COG	No COG association	NC	-
PANA_2499	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0438
PANA_2500	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0463
PANA_2501	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0463
PANA_2502	NO COG	No COG association	NC	-
PANA_2503	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3206
PANA_2504	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0394
PANA_2505	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1596
PANA_2506	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2148
PANA_2507	METABOLISM	Inorganic ion transport and metabolism	P	COG0861
PANA_2508	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2982

PANA_2509	METABOLISM	Nucleotide transport and metabolism	F	COG0717
PANA_2510	METABOLISM	Nucleotide transport and metabolism	F	COG0572
PANA_2511	NO COG	No COG association	NC	-
PANA_2512	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2199
PANA_2513	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0443
PANA_2515	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0845
PANA_2516	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0841
PANA_2517	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0841
PANA_2518	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_2519	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_2520	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_2520	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_2521	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0826
PANA_2522	METABOLISM	Lipid transport and metabolism	I	COG1597
PANA_2522	POORLY CHARACTERIZED	General function prediction only	R	COG1597
PANA_2523	METABOLISM	Carbohydrate transport and metabolism	G	COG0246
PANA_2524	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1802
PANA_2525	METABOLISM	Coenzyme transport and metabolism	H	COG0351
PANA_2526	METABOLISM	Coenzyme transport and metabolism	H	COG2145
PANA_2528	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG3634
PANA_2529	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0450
PANA_2531	POORLY CHARACTERIZED	General function prediction only	R	COG0456
PANA_2532	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0489
PANA_2533	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0143
PANA_2534	POORLY CHARACTERIZED	General function prediction only	R	COG1380
PANA_2535	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1346
PANA_2536	METABOLISM	Nucleotide transport and metabolism	F	COG0295
PANA_2537	METABOLISM	Energy production and conversion	C	COG0281
PANA_2538	POORLY CHARACTERIZED	Function unknown	S	COG2949
PANA_2539	METABOLISM	Carbohydrate transport and metabolism	G	COG4211
PANA_2540	METABOLISM	Carbohydrate transport and metabolism	G	COG1129
PANA_2541	METABOLISM	Carbohydrate transport and metabolism	G	COG1879
PANA_2542	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_2543	POORLY CHARACTERIZED	Function unknown	S	COG2311
PANA_2544	METABOLISM	Coenzyme transport and metabolism	H	COG0302
PANA_2545	NO COG	No COG association	NC	-
PANA_2546	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_2547	METABOLISM	Energy production and conversion	C	COG1062
PANA_2548	POORLY CHARACTERIZED	General function prediction only	R	COG0627
PANA_2549	METABOLISM	Inorganic ion transport and metabolism	P	COG1120
PANA_2549	METABOLISM	Coenzyme transport and metabolism	H	COG1120
PANA_2550	METABOLISM	Inorganic ion transport and metabolism	P	COG0609
PANA_2551	METABOLISM	Amino acid transport and metabolism	E	COG0833
PANA_2552	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_2554	POORLY CHARACTERIZED	Function unknown	S	COG2855
PANA_2556	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0648
PANA_2557	METABOLISM	Carbohydrate transport and metabolism	G	COG1299
PANA_2558	METABOLISM	Carbohydrate transport and metabolism	G	COG1105
PANA_2559	METABOLISM	Carbohydrate transport and metabolism	G	COG4668
PANA_2565	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0231
PANA_2566	POORLY CHARACTERIZED	General function prediction only	R	COG0523
PANA_2567	METABOLISM	Lipid transport and metabolism	I	COG0671
PANA_2568	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0791
PANA_2569	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2200
PANA_2570	METABOLISM	Amino acid transport and metabolism	E	COG4166
PANA_2571	POORLY CHARACTERIZED	General function prediction only	R	COG4174
PANA_2572	POORLY CHARACTERIZED	General function prediction only	R	COG4239
PANA_2573	POORLY CHARACTERIZED	General function prediction only	R	COG4172
PANA_2574	NO COG	No COG association	NC	-
PANA_2575	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_2576	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1187
PANA_2577	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1061
PANA_2577	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1061
PANA_2578	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1825

PANA_2579	POORLY CHARACTERIZED	General function prediction only	R	COG3081
PANA_2580	POORLY CHARACTERIZED	Function unknown	S	COG3082
PANA_2581	POORLY CHARACTERIZED	General function prediction only	R	COG3083
PANA_2598	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG4615
PANA_2598	METABOLISM	Inorganic ion transport and metabolism	P	COG4615
PANA_2599	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG3145
PANA_2600	METABOLISM	Coenzyme transport and metabolism	H	COG1477
PANA_2602	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3203
PANA_2603	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1566
PANA_2604	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG2274
PANA_2608	NO COG	No COG association	NC	-
PANA_2609	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_2610	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2197
PANA_2610	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2197
PANA_2611	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_2612	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0188
PANA_2615	METABOLISM	Coenzyme transport and metabolism	H	COG2227
PANA_2616	METABOLISM	Nucleotide transport and metabolism	F	COG0209
PANA_2617	METABOLISM	Nucleotide transport and metabolism	F	COG0208
PANA_2618	METABOLISM	Energy production and conversion	C	COG0633
PANA_2619	POORLY CHARACTERIZED	General function prediction only	R	COG1058
PANA_2620	POORLY CHARACTERIZED	Function unknown	S	COG4575
PANA_2621	METABOLISM	Inorganic ion transport and metabolism	P	COG3131
PANA_2622	POORLY CHARACTERIZED	General function prediction only	R	COG2153
PANA_2624	METABOLISM	Energy production and conversion	C	COG1007
PANA_2625	METABOLISM	Energy production and conversion	C	COG1008
PANA_2626	METABOLISM	Energy production and conversion	C	COG1009
PANA_2626	METABOLISM	Inorganic ion transport and metabolism	P	COG1009
PANA_2627	METABOLISM	Energy production and conversion	C	COG0713
PANA_2628	METABOLISM	Energy production and conversion	C	COG0839
PANA_2629	METABOLISM	Energy production and conversion	C	COG1143
PANA_2630	METABOLISM	Energy production and conversion	C	COG1005
PANA_2631	METABOLISM	Energy production and conversion	C	COG1034
PANA_2632	METABOLISM	Energy production and conversion	C	COG1894
PANA_2633	METABOLISM	Energy production and conversion	C	COG1905
PANA_2634	METABOLISM	Energy production and conversion	C	COG0649
PANA_2635	METABOLISM	Energy production and conversion	C	COG0377
PANA_2636	METABOLISM	Energy production and conversion	C	COG0838
PANA_2637	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_2638	METABOLISM	Amino acid transport and metabolism	E	COG0436
PANA_2639	POORLY CHARACTERIZED	General function prediction only	R	COG1896
PANA_2640	POORLY CHARACTERIZED	General function prediction only	R	COG0637
PANA_2641	POORLY CHARACTERIZED	Function unknown	S	COG3013
PANA_2643	POORLY CHARACTERIZED	Function unknown	S	COG3092
PANA_2644	METABOLISM	Energy production and conversion	C	COG0282
PANA_2645	METABOLISM	Energy production and conversion	C	COG0280
PANA_2647	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1167
PANA_2647	METABOLISM	Amino acid transport and metabolism	E	COG1167
PANA_2648	METABOLISM	Amino acid transport and metabolism	E	COG0160
PANA_2652	NO COG	No COG association	NC	-
PANA_2653	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0625
PANA_2654	POORLY CHARACTERIZED	General function prediction only	R	COG1090
PANA_2659	METABOLISM	Amino acid transport and metabolism	E	COG4598
PANA_2660	METABOLISM	Amino acid transport and metabolism	E	COG4160
PANA_2661	METABOLISM	Amino acid transport and metabolism	E	COG4215
PANA_2662	METABOLISM	Amino acid transport and metabolism	E	COG0834
PANA_2662	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0834
PANA_2663	METABOLISM	Coenzyme transport and metabolism	H	COG0163
PANA_2664	METABOLISM	Nucleotide transport and metabolism	F	COG0034
PANA_2665	POORLY CHARACTERIZED	General function prediction only	R	COG1286
PANA_2667	METABOLISM	Coenzyme transport and metabolism	H	COG0285
PANA_2668	METABOLISM	Lipid transport and metabolism	I	COG0777
PANA_2669	POORLY CHARACTERIZED	Function unknown	S	COG0586
PANA_2670	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0101

PANA_2671	METABOLISM	Amino acid transport and metabolism	E	COG0136
PANA_2672	METABOLISM	Coenzyme transport and metabolism	H	COG0111
PANA_2672	METABOLISM	Amino acid transport and metabolism	E	COG0111
PANA_2673	NO COG	No COG association	NC	-
PANA_2679	NO COG	No COG association	NC	-
PANA_2681	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1167
PANA_2681	METABOLISM	Amino acid transport and metabolism	E	COG1167
PANA_2682	METABOLISM	Lipid transport and metabolism	I	COG0304
PANA_2682	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG0304
PANA_2683	POORLY CHARACTERIZED	Function unknown	S	COG4121
PANA_2684	NO COG	No COG association	NC	-
PANA_2685	POORLY CHARACTERIZED	Function unknown	S	COG3101
PANA_2686	POORLY CHARACTERIZED	General function prediction only	R	COG0730
PANA_2687	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3770
PANA_2688	METABOLISM	Amino acid transport and metabolism	E	COG0082
PANA_2689	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2890
PANA_2690	POORLY CHARACTERIZED	Function unknown	S	COG2840
PANA_2692	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2062
PANA_2693	METABOLISM	Lipid transport and metabolism	I	COG1250
PANA_2694	METABOLISM	Lipid transport and metabolism	I	COG0183
PANA_2695	POORLY CHARACTERIZED	Function unknown	S	COG3691
PANA_2696	METABOLISM	Lipid transport and metabolism	I	COG2067
PANA_2697	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2853
PANA_2698	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG3088
PANA_2699	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0526
PANA_2699	METABOLISM	Energy production and conversion	C	COG0526
PANA_2700	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1138
PANA_2701	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG2332
PANA_2702	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG3114
PANA_2703	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0755
PANA_2704	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG2386
PANA_2705	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG4133
PANA_2706	METABOLISM	Inorganic ion transport and metabolism	P	COG2116
PANA_2731	NO COG	No COG association	NC	-
PANA_2732	METABOLISM	Carbohydrate transport and metabolism	G	COG3386
PANA_2734	NO COG	No COG association	NC	-
PANA_2736	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1414
PANA_2737	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_2738	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1566
PANA_2740	METABOLISM	Amino acid transport and metabolism	E	COG0436
PANA_2741	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1670
PANA_2742	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3275
PANA_2743	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3279
PANA_2743	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3279
PANA_2744	METABOLISM	Carbohydrate transport and metabolism	G	COG0837
PANA_2745	METABOLISM	Carbohydrate transport and metabolism	G	COG3961
PANA_2745	METABOLISM	Coenzyme transport and metabolism	H	COG3961
PANA_2745	POORLY CHARACTERIZED	General function prediction only	R	COG3961
PANA_2746	METABOLISM	Energy production and conversion	C	COG0667
PANA_2747	NO COG	No COG association	NC	-
PANA_2748	POORLY CHARACTERIZED	General function prediction only	R	COG4122
PANA_2749	METABOLISM	Inorganic ion transport and metabolism	P	COG1914
PANA_2750	METABOLISM	Nucleotide transport and metabolism	F	COG1972
PANA_2752	METABOLISM	Inorganic ion transport and metabolism	P	COG2116
PANA_2753	NO COG	No COG association	NC	-
PANA_2754	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0008
PANA_2755	NO COG	No COG association	NC	-
PANA_2756	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_2757	POORLY CHARACTERIZED	General function prediction only	R	COG0385
PANA_2758	POORLY CHARACTERIZED	Function unknown	S	COG3530
PANA_2759	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0272
PANA_2760	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG3115
PANA_2761	METABOLISM	Amino acid transport and metabolism	E	COG2981
PANA_2762	METABOLISM	Amino acid transport and metabolism	E	COG0031

PANA_2763	METABOLISM	Carbohydrate transport and metabolism	G	COG1925
PANA_2764	METABOLISM	Carbohydrate transport and metabolism	G	COG1080
PANA_2765	METABOLISM	Carbohydrate transport and metabolism	G	COG2190
PANA_2766	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_2767	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_2767	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_2768	METABOLISM	Amino acid transport and metabolism	E	COG0031
PANA_2769	METABOLISM	Inorganic ion transport and metabolism	P	COG1118
PANA_2770	METABOLISM	Inorganic ion transport and metabolism	P	COG4208
PANA_2771	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0555
PANA_2772	METABOLISM	Inorganic ion transport and metabolism	P	COG4150
PANA_2773	NO COG	No COG association	NC	-
PANA_2774	NO COG	No COG association	NC	-
PANA_2775	POORLY CHARACTERIZED	General function prediction only	R	COG0456
PANA_2776	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0860
PANA_2777	METABOLISM	Coenzyme transport and metabolism	H	COG0408
PANA_2779	METABOLISM	Carbohydrate transport and metabolism	G	COG0176
PANA_2780	METABOLISM	Carbohydrate transport and metabolism	G	COG0021
PANA_2784	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1680
PANA_2785	NO COG	No COG association	NC	-
PANA_2793	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0494
PANA_2793	POORLY CHARACTERIZED	General function prediction only	R	COG0494
PANA_2794	METABOLISM	Amino acid transport and metabolism	E	COG1113
PANA_2795	NO COG	No COG association	NC	-
PANA_2796	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3850
PANA_2797	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2197
PANA_2797	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2197
PANA_2798	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG0841
PANA_2799	METABOLISM	Inorganic ion transport and metabolism	P	COG1393
PANA_2800	METABOLISM	Amino acid transport and metabolism	E	COG0624
PANA_2802	POORLY CHARACTERIZED	General function prediction only	R	COG0400
PANA_2803	POORLY CHARACTERIZED	General function prediction only	R	COG1444
PANA_2804	POORLY CHARACTERIZED	General function prediction only	R	COG2321
PANA_2805	METABOLISM	Nucleotide transport and metabolism	F	COG0152
PANA_2806	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3317
PANA_2807	METABOLISM	Amino acid transport and metabolism	E	COG0329
PANA_2807	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0329
PANA_2808	METABOLISM	Amino acid transport and metabolism	E	COG2716
PANA_2809	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1225
PANA_2815	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3706
PANA_2816	POORLY CHARACTERIZED	General function prediction only	R	COG1409
PANA_2817	METABOLISM	Amino acid transport and metabolism	E	COG3842
PANA_2818	METABOLISM	Amino acid transport and metabolism	E	COG1177
PANA_2819	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0555
PANA_2820	METABOLISM	Inorganic ion transport and metabolism	P	COG1840
PANA_2821	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_2822	POORLY CHARACTERIZED	General function prediction only	R	COG4783
PANA_2823	METABOLISM	Inorganic ion transport and metabolism	P	COG1393
PANA_2824	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0593
PANA_2825	METABOLISM	Nucleotide transport and metabolism	F	COG2233
PANA_2826	METABOLISM	Nucleotide transport and metabolism	F	COG0035
PANA_2827	METABOLISM	Nucleotide transport and metabolism	F	COG0150
PANA_2828	METABOLISM	Nucleotide transport and metabolism	F	COG0299
PANA_2829	METABOLISM	Inorganic ion transport and metabolism	P	COG1117
PANA_2830	METABOLISM	Inorganic ion transport and metabolism	P	COG4985
PANA_2831	POORLY CHARACTERIZED	General function prediction only	R	COG4590
PANA_2832	METABOLISM	Inorganic ion transport and metabolism	P	COG0855
PANA_2833	METABOLISM	Nucleotide transport and metabolism	F	COG0248
PANA_2833	METABOLISM	Inorganic ion transport and metabolism	P	COG0248
PANA_2834	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG0500
PANA_2834	POORLY CHARACTERIZED	General function prediction only	R	COG0500
PANA_2836	METABOLISM	Inorganic ion transport and metabolism	P	COG2239
PANA_2840	NO COG	No COG association	NC	-
PANA_2848	METABOLISM	Nucleotide transport and metabolism	F	COG0519

PANA_2849	METABOLISM	Nucleotide transport and metabolism	F	COG0516
PANA_2850	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1570
PANA_2851	METABOLISM	Amino acid transport and metabolism	E	COG3227
PANA_2853	METABOLISM	Carbohydrate transport and metabolism	G	COG4993
PANA_2854	POORLY CHARACTERIZED	General function prediction only	R	COG1160
PANA_2855	POORLY CHARACTERIZED	Function unknown	S	COG1520
PANA_2856	POORLY CHARACTERIZED	Function unknown	S	COG2976
PANA_2857	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0124
PANA_2858	METABOLISM	Lipid transport and metabolism	I	COG0821
PANA_2859	POORLY CHARACTERIZED	Function unknown	S	COG1426
PANA_2860	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG3063
PANA_2860	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG3063
PANA_2861	POORLY CHARACTERIZED	General function prediction only	R	COG0820
PANA_2862	METABOLISM	Nucleotide transport and metabolism	F	COG0105
PANA_2865	METABOLISM	Inorganic ion transport and metabolism	P	COG2897
PANA_2866	NO COG	No COG association	NC	-
PANA_2867	METABOLISM	Amino acid transport and metabolism	E	COG0260
PANA_2868	METABOLISM	Amino acid transport and metabolism	E	COG1104
PANA_2869	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1959
PANA_2870	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0565
PANA_2871	METABOLISM	Carbohydrate transport and metabolism	G	COG0483
PANA_2873	POORLY CHARACTERIZED	General function prediction only	R	COG2215
PANA_2874	POORLY CHARACTERIZED	General function prediction only	R	COG3683
PANA_2875	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3711
PANA_2876	NO COG	No COG association	NC	-
PANA_2877	METABOLISM	Amino acid transport and metabolism	E	COG0112
PANA_2878	METABOLISM	Energy production and conversion	C	COG1018
PANA_2879	METABOLISM	Amino acid transport and metabolism	E	COG0347
PANA_2880	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2204
PANA_2881	NO COG	No COG association	NC	-
PANA_2882	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_2883	METABOLISM	Nucleotide transport and metabolism	F	COG0046
PANA_2884	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG4623
PANA_2885	METABOLISM	Nucleotide transport and metabolism	F	COG0590
PANA_2885	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0590
PANA_2886	METABOLISM	Amino acid transport and metabolism	E	COG0560
PANA_2887	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1737
PANA_2888	METABOLISM	Energy production and conversion	C	COG1143
PANA_2889	METABOLISM	Lipid transport and metabolism	I	COG0736
PANA_2890	METABOLISM	Coenzyme transport and metabolism	H	COG0854
PANA_2891	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1381
PANA_2892	POORLY CHARACTERIZED	General function prediction only	R	COG1159
PANA_2893	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0571
PANA_2894	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0681
PANA_2895	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0481
PANA_2896	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3086
PANA_2897	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3026
PANA_2898	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3073
PANA_2899	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1595
PANA_2900	METABOLISM	Coenzyme transport and metabolism	H	COG0029
PANA_2901	POORLY CHARACTERIZED	General function prediction only	R	COG4123
PANA_2902	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0513
PANA_2902	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0513
PANA_2902	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0513
PANA_2903	POORLY CHARACTERIZED	General function prediction only	R	COG3445
PANA_2904	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0692
PANA_2905	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0576
PANA_2906	METABOLISM	Carbohydrate transport and metabolism	G	COG0061
PANA_2907	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0497
PANA_2908	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2913
PANA_2909	POORLY CHARACTERIZED	Function unknown	S	COG2914
PANA_2910	METABOLISM	Lipid transport and metabolism	I	COG2867
PANA_2911	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0691
PANA_2944	POORLY CHARACTERIZED	General function prediction only	R	COG0596

PANA_2955	METABOLISM	Carbohydrate transport and metabolism	G	COG2271
PANA_2957	METABOLISM	Energy production and conversion	C	COG0604
PANA_2957	POORLY CHARACTERIZED	General function prediction only	R	COG0604
PANA_2973	METABOLISM	Inorganic ion transport and metabolism	P	COG1120
PANA_2973	METABOLISM	Coenzyme transport and metabolism	H	COG1120
PANA_2977	METABOLISM	Inorganic ion transport and metabolism	P	COG0474
PANA_2978	METABOLISM	Amino acid transport and metabolism	E	COG0747
PANA_2979	METABOLISM	Amino acid transport and metabolism	E	COG0601
PANA_2979	METABOLISM	Inorganic ion transport and metabolism	P	COG0601
PANA_2980	METABOLISM	Amino acid transport and metabolism	E	COG1173
PANA_2980	METABOLISM	Inorganic ion transport and metabolism	P	COG1173
PANA_2981	METABOLISM	Amino acid transport and metabolism	E	COG0444
PANA_2981	METABOLISM	Inorganic ion transport and metabolism	P	COG0444
PANA_2982	METABOLISM	Amino acid transport and metabolism	E	COG1124
PANA_2982	METABOLISM	Inorganic ion transport and metabolism	P	COG1124
PANA_2983	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1670
PANA_2985	METABOLISM	Amino acid transport and metabolism	E	COG2040
PANA_2986	METABOLISM	Amino acid transport and metabolism	E	COG0833
PANA_2987	NO COG	No COG association	NC	-
PANA_2988	POORLY CHARACTERIZED	Function unknown	S	COG4575
PANA_2989	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0695
PANA_2990	METABOLISM	Nucleotide transport and metabolism	F	COG1780
PANA_2991	METABOLISM	Nucleotide transport and metabolism	F	COG0209
PANA_2992	METABOLISM	Nucleotide transport and metabolism	F	COG0208
PANA_2996	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_2997	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1846
PANA_2998	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1566
PANA_2999	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_3000	POORLY CHARACTERIZED	General function prediction only	R	COG1064
PANA_3001	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1733
PANA_3002	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0566
PANA_3003	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG3118
PANA_3004	POORLY CHARACTERIZED	Function unknown	S	COG3148
PANA_3005	METABOLISM	Energy production and conversion	C	COG1042
PANA_3006	METABOLISM	Lipid transport and metabolism	I	COG1502
PANA_3007	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_3010	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0542
PANA_3011	POORLY CHARACTERIZED	Function unknown	S	COG1496
PANA_3012	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0564
PANA_3013	POORLY CHARACTERIZED	General function prediction only	R	COG4105
PANA_3014	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1544
PANA_3015	METABOLISM	Amino acid transport and metabolism	E	COG0077
PANA_3016	METABOLISM	Amino acid transport and metabolism	E	COG0287
PANA_3017	METABOLISM	Amino acid transport and metabolism	E	COG0722
PANA_3018	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1670
PANA_3020	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0335
PANA_3021	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0336
PANA_3022	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0806
PANA_3023	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0228
PANA_3024	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0541
PANA_3025	POORLY CHARACTERIZED	General function prediction only	R	COG4137
PANA_3026	METABOLISM	Inorganic ion transport and metabolism	P	COG4536
PANA_3027	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1854
PANA_3028	METABOLISM	Coenzyme transport and metabolism	H	COG2918
PANA_3030	POORLY CHARACTERIZED	General function prediction only	R	COG0637
PANA_3031	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1551
PANA_3032	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0013
PANA_3033	POORLY CHARACTERIZED	General function prediction only	R	COG2137
PANA_3034	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0468
PANA_3035	POORLY CHARACTERIZED	General function prediction only	R	COG1546
PANA_3036	POORLY CHARACTERIZED	General function prediction only	R	COG4106
PANA_3037	METABOLISM	Inorganic ion transport and metabolism	P	COG2824
PANA_3038	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2951
PANA_3039	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0249

PANA_3041	METABOLISM	Coenzyme transport and metabolism	H	COG2045
PANA_3041	POORLY CHARACTERIZED	General function prediction only	R	COG2045
PANA_3042	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0568
PANA_3043	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0739
PANA_3044	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG2518
PANA_3045	POORLY CHARACTERIZED	General function prediction only	R	COG0496
PANA_3046	POORLY CHARACTERIZED	Function unknown	S	COG0585
PANA_3047	METABOLISM	Lipid transport and metabolism	I	COG0245
PANA_3048	METABOLISM	Lipid transport and metabolism	I	COG1211
PANA_3049	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG2919
PANA_3050	NO COG	No COG association	NC	-
PANA_3051	METABOLISM	Inorganic ion transport and metabolism	P	COG0529
PANA_3052	METABOLISM	Inorganic ion transport and metabolism	P	COG2895
PANA_3053	METABOLISM	Amino acid transport and metabolism	E	COG0175
PANA_3053	METABOLISM	Coenzyme transport and metabolism	H	COG0175
PANA_3054	METABOLISM	Coenzyme transport and metabolism	H	COG0007
PANA_3055	POORLY CHARACTERIZED	General function prediction only	R	COG2234
PANA_3064	METABOLISM	Amino acid transport and metabolism	E	COG0175
PANA_3064	METABOLISM	Coenzyme transport and metabolism	H	COG0175
PANA_3065	METABOLISM	Inorganic ion transport and metabolism	P	COG0155
PANA_3066	METABOLISM	Inorganic ion transport and metabolism	P	COG0369
PANA_3068	METABOLISM	Coenzyme transport and metabolism	H	COG0720
PANA_3070	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0602
PANA_3075	METABOLISM	Carbohydrate transport and metabolism	G	COG0148
PANA_3076	METABOLISM	Nucleotide transport and metabolism	F	COG0504
PANA_3077	POORLY CHARACTERIZED	General function prediction only	R	COG3956
PANA_3078	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0317
PANA_3078	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0317
PANA_3079	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2265
PANA_3080	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_3081	METABOLISM	Carbohydrate transport and metabolism	G	COG1929
PANA_3082	METABOLISM	Lipid transport and metabolism	I	COG2084
PANA_3083	METABOLISM	Carbohydrate transport and metabolism	G	COG3836
PANA_3084	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG4948
PANA_3084	POORLY CHARACTERIZED	General function prediction only	R	COG4948
PANA_3085	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG4948
PANA_3085	POORLY CHARACTERIZED	General function prediction only	R	COG4948
PANA_3086	METABOLISM	Carbohydrate transport and metabolism	G	COG2271
PANA_3087	METABOLISM	Carbohydrate transport and metabolism	G	COG2721
PANA_3088	METABOLISM	Energy production and conversion	C	COG0716
PANA_3089	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0564
PANA_3090	POORLY CHARACTERIZED	Function unknown	S	COG3098
PANA_3091	NO COG	No COG association	NC	-
PANA_3092	POORLY CHARACTERIZED	Function unknown	S	COG2904
PANA_3093	POORLY CHARACTERIZED	General function prediction only	R	COG1611
PANA_3094	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0258
PANA_3095	POORLY CHARACTERIZED	General function prediction only	R	COG2933
PANA_3096	POORLY CHARACTERIZED	Function unknown	S	COG2363
PANA_3097	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3098	NO COG	No COG association	NC	-
PANA_3099	METABOLISM	Amino acid transport and metabolism	E	COG0520
PANA_3100	POORLY CHARACTERIZED	General function prediction only	R	COG2166
PANA_3101	METABOLISM	Coenzyme transport and metabolism	H	COG1179
PANA_3102	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2821
PANA_3104	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0860
PANA_3105	METABOLISM	Amino acid transport and metabolism	E	COG0548
PANA_3106	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0507
PANA_3107	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1074
PANA_3108	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1025
PANA_3109	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1330
PANA_3110	NO COG	No COG association	NC	-
PANA_3112	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG4795
PANA_3113	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG4970
PANA_3113	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG4970

PANA_3114	METABOLISM	Nucleotide transport and metabolism	F	COG0207
PANA_3115	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0682
PANA_3116	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3605
PANA_3117	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0494
PANA_3117	POORLY CHARACTERIZED	General function prediction only	R	COG0494
PANA_3118	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG3066
PANA_3119	METABOLISM	Inorganic ion transport and metabolism	P	COG0861
PANA_3120	NO COG	No COG association	NC	-
PANA_3121	METABOLISM	Energy production and conversion	C	COG0667
PANA_3122	METABOLISM	Amino acid transport and metabolism	E	COG0019
PANA_3123	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3124	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_3125	NO COG	No COG association	NC	-
PANA_3127	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1190
PANA_3128	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1186
PANA_3129	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0608
PANA_3130	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1651
PANA_3131	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG4974
PANA_3179	METABOLISM	Energy production and conversion	C	COG0716
PANA_3180	NO COG	No COG association	NC	-
PANA_3181	POORLY CHARACTERIZED	Function unknown	S	COG2938
PANA_3182	POORLY CHARACTERIZED	General function prediction only	R	COG0354
PANA_3183	POORLY CHARACTERIZED	General function prediction only	R	COG1272
PANA_3185	METABOLISM	Carbohydrate transport and metabolism	G	COG2723
PANA_3186	METABOLISM	Lipid transport and metabolism	I	COG1028
PANA_3186	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1028
PANA_3186	POORLY CHARACTERIZED	General function prediction only	R	COG1028
PANA_3187	METABOLISM	Amino acid transport and metabolism	E	COG1003
PANA_3188	METABOLISM	Amino acid transport and metabolism	E	COG0509
PANA_3189	METABOLISM	Amino acid transport and metabolism	E	COG0404
PANA_3190	METABOLISM	Coenzyme transport and metabolism	H	COG0654
PANA_3190	METABOLISM	Energy production and conversion	C	COG0654
PANA_3191	METABOLISM	Coenzyme transport and metabolism	H	COG0654
PANA_3191	METABOLISM	Energy production and conversion	C	COG0654
PANA_3192	METABOLISM	Amino acid transport and metabolism	E	COG0006
PANA_3193	POORLY CHARACTERIZED	Function unknown	S	COG3079
PANA_3194	POORLY CHARACTERIZED	Function unknown	S	COG3027
PANA_3195	METABOLISM	Coenzyme transport and metabolism	H	COG0212
PANA_3196	METABOLISM	Coenzyme transport and metabolism	H	COG0111
PANA_3196	METABOLISM	Amino acid transport and metabolism	E	COG0111
PANA_3197	METABOLISM	Carbohydrate transport and metabolism	G	COG0120
PANA_3198	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3199	POORLY CHARACTERIZED	Function unknown	S	COG2968
PANA_3200	POORLY CHARACTERIZED	General function prediction only	R	COG1279
PANA_3201	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0668
PANA_3203	METABOLISM	Carbohydrate transport and metabolism	G	COG0191
PANA_3204	METABOLISM	Carbohydrate transport and metabolism	G	COG0126
PANA_3205	METABOLISM	Carbohydrate transport and metabolism	G	COG0057
PANA_3206	METABOLISM	Carbohydrate transport and metabolism	G	COG0021
PANA_3207	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0501
PANA_3208	METABOLISM	Amino acid transport and metabolism	E	COG0010
PANA_3209	METABOLISM	Amino acid transport and metabolism	E	COG1166
PANA_3210	METABOLISM	Coenzyme transport and metabolism	H	COG0192
PANA_3211	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_3213	POORLY CHARACTERIZED	Function unknown	S	COG3091
PANA_3214	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2356
PANA_3215	POORLY CHARACTERIZED	Function unknown	S	COG1385
PANA_3216	METABOLISM	Coenzyme transport and metabolism	H	COG0189
PANA_3216	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0189
PANA_3217	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1678
PANA_3218	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0816
PANA_3219	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG2805
PANA_3219	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG2805
PANA_3220	POORLY CHARACTERIZED	General function prediction only	R	COG0325

PANA_3221	POORLY CHARACTERIZED	Function unknown	S	COG0762
PANA_3222	METABOLISM	Nucleotide transport and metabolism	F	COG0127
PANA_3223	METABOLISM	Coenzyme transport and metabolism	H	COG0635
PANA_3224	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2188
PANA_3225	METABOLISM	Carbohydrate transport and metabolism	G	COG1762
PANA_3225	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1762
PANA_3228	POORLY CHARACTERIZED	Function unknown	S	COG3037
PANA_3229	NO COG	No COG association	NC	-
PANA_3230	POORLY CHARACTERIZED	Function unknown	S	COG3171
PANA_3231	POORLY CHARACTERIZED	General function prediction only	R	COG0220
PANA_3235	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1194
PANA_3236	METABOLISM	Energy production and conversion	C	COG2924
PANA_3236	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG2924
PANA_3237	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0741
PANA_3238	METABOLISM	Amino acid transport and metabolism	E	COG1982
PANA_3239	METABOLISM	Energy production and conversion	C	COG0604
PANA_3239	POORLY CHARACTERIZED	General function prediction only	R	COG0604
PANA_3241	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1695
PANA_3242	METABOLISM	Inorganic ion transport and metabolism	P	COG2375
PANA_3243	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_3245	METABOLISM	Energy production and conversion	C	COG1294
PANA_3246	METABOLISM	Energy production and conversion	C	COG1271
PANA_3247	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1733
PANA_3248	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0702
PANA_3248	METABOLISM	Carbohydrate transport and metabolism	G	COG0702
PANA_3251	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0640
PANA_3253	METABOLISM	Carbohydrate transport and metabolism	G	COG1626
PANA_3256	NO COG	No COG association	NC	-
PANA_3257	POORLY CHARACTERIZED	General function prediction only	R	COG0596
PANA_3258	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2199
PANA_3268	METABOLISM	Carbohydrate transport and metabolism	G	COG4130
PANA_3274	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3275	METABOLISM	Amino acid transport and metabolism	E	COG0624
PANA_3276	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_3297	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0768
PANA_3298	METABOLISM	Nucleotide transport and metabolism	F	COG0813
PANA_3299	METABOLISM	Carbohydrate transport and metabolism	G	COG1015
PANA_3300	METABOLISM	Nucleotide transport and metabolism	F	COG0213
PANA_3301	METABOLISM	Nucleotide transport and metabolism	F	COG0274
PANA_3302	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0084
PANA_3303	POORLY CHARACTERIZED	General function prediction only	R	COG4667
PANA_3305	POORLY CHARACTERIZED	General function prediction only	R	COG2823
PANA_3306	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG4108
PANA_3308	METABOLISM	Inorganic ion transport and metabolism	P	COG1528
PANA_3317	POORLY CHARACTERIZED	Function unknown	S	COG4297
PANA_3319	METABOLISM	Amino acid transport and metabolism	E	COG1126
PANA_3320	METABOLISM	Amino acid transport and metabolism	E	COG0765
PANA_3321	POORLY CHARACTERIZED	General function prediction only	R	COG3393
PANA_3323	METABOLISM	Energy production and conversion	C	COG2141
PANA_3324	POORLY CHARACTERIZED	General function prediction only	R	COG1473
PANA_3325	METABOLISM	Energy production and conversion	C	COG2141
PANA_3327	POORLY CHARACTERIZED	General function prediction only	R	COG1473
PANA_3328	NO COG	No COG association	NC	-
PANA_3329	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1167
PANA_3329	METABOLISM	Amino acid transport and metabolism	E	COG1167
PANA_3330	POORLY CHARACTERIZED	Function unknown	S	COG5430
PANA_3331	POORLY CHARACTERIZED	Function unknown	S	COG5430
PANA_3332	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG3121
PANA_3332	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG3121
PANA_3333	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG3188
PANA_3333	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG3188
PANA_3334	POORLY CHARACTERIZED	Function unknown	S	COG5430
PANA_3335	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_3336	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0848

PANA_3337	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0811
PANA_3352	METABOLISM	Amino acid transport and metabolism	E	COG0626
PANA_3353	POORLY CHARACTERIZED	Function unknown	S	COG0586
PANA_3354	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_3355	POORLY CHARACTERIZED	General function prediction only	R	COG0656
PANA_3357	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG2132
PANA_3358	METABOLISM	Lipid transport and metabolism	I	COG0204
PANA_3359	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0188
PANA_3365	METABOLISM	Amino acid transport and metabolism	E	COG0834
PANA_3365	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0834
PANA_3366	POORLY CHARACTERIZED	General function prediction only	R	COG2249
PANA_3367	POORLY CHARACTERIZED	Function unknown	S	COG1359
PANA_3368	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0187
PANA_3369	POORLY CHARACTERIZED	General function prediction only	R	COG3150
PANA_3370	POORLY CHARACTERIZED	General function prediction only	R	COG1409
PANA_3371	POORLY CHARACTERIZED	Function unknown	S	COG3151
PANA_3372	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2816
PANA_3373	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1538
PANA_3373	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1538
PANA_3374	POORLY CHARACTERIZED	Function unknown	S	COG5463
PANA_3375	METABOLISM	Amino acid transport and metabolism	E	COG0754
PANA_3376	POORLY CHARACTERIZED	Function unknown	S	COG3384
PANA_3377	METABOLISM	Coenzyme transport and metabolism	H	COG0108
PANA_3378	POORLY CHARACTERIZED	Function unknown	S	COG2960
PANA_3379	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2870
PANA_3380	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1391
PANA_3380	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1391
PANA_3381	POORLY CHARACTERIZED	Function unknown	S	COG3025
PANA_3382	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3103
PANA_3383	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0617
PANA_3384	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1968
PANA_3385	METABOLISM	Coenzyme transport and metabolism	H	COG1539
PANA_3386	POORLY CHARACTERIZED	Function unknown	S	COG0344
PANA_3387	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0533
PANA_3388	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0828
PANA_3389	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0358
PANA_3390	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0568
PANA_3391	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG3663
PANA_3397	POORLY CHARACTERIZED	General function prediction only	R	COG3499
PANA_3402	POORLY CHARACTERIZED	General function prediction only	R	COG3498
PANA_3403	POORLY CHARACTERIZED	General function prediction only	R	COG3497
PANA_3407	POORLY CHARACTERIZED	General function prediction only	R	COG4385
PANA_3408	POORLY CHARACTERIZED	General function prediction only	R	COG3948
PANA_3409	POORLY CHARACTERIZED	General function prediction only	R	COG3628
PANA_3410	POORLY CHARACTERIZED	General function prediction only	R	COG4540
PANA_3412	NO COG	No COG association	NC	-
PANA_3414	POORLY CHARACTERIZED	General function prediction only	R	COG3772
PANA_3423	METABOLISM	Carbohydrate transport and metabolism	G	COG0191
PANA_3424	METABOLISM	Carbohydrate transport and metabolism	G	COG1299
PANA_3425	METABOLISM	Carbohydrate transport and metabolism	G	COG1762
PANA_3425	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1762
PANA_3426	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3711
PANA_3439	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2200
PANA_3442	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1181
PANA_3443	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2199
PANA_3444	METABOLISM	Inorganic ion transport and metabolism	P	COG0659
PANA_3445	METABOLISM	Inorganic ion transport and metabolism	P	COG0288
PANA_3446	POORLY CHARACTERIZED	General function prediction only	R	COG2503
PANA_3447	POORLY CHARACTERIZED	Function unknown	S	COG1496
PANA_3448	METABOLISM	Energy production and conversion	C	COG1902
PANA_3449	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2813
PANA_3450	POORLY CHARACTERIZED	General function prediction only	R	COG1451
PANA_3451	METABOLISM	Inorganic ion transport and metabolism	P	COG0038
PANA_3452	POORLY CHARACTERIZED	General function prediction only	R	COG0673

PANA_3453	METABOLISM	Inorganic ion transport and metabolism	P	COG0861
PANA_3454	METABOLISM	Carbohydrate transport and metabolism	G	COG2721
PANA_3455	METABOLISM	Carbohydrate transport and metabolism	G	COG0246
PANA_3456	METABOLISM	Carbohydrate transport and metabolism	G	COG1904
PANA_3457	METABOLISM	Carbohydrate transport and metabolism	G	COG2271
PANA_3458	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2186
PANA_3459	POORLY CHARACTERIZED	Function unknown	S	COG0586
PANA_3460	NO COG	No COG association	NC	-
PANA_3461	NO COG	No COG association	NC	-
PANA_3463	POORLY CHARACTERIZED	Function unknown	S	COG5393
PANA_3464	NO COG	No COG association	NC	-
PANA_3465	POORLY CHARACTERIZED	Function unknown	S	COG2259
PANA_3469	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0435
PANA_3473	POORLY CHARACTERIZED	General function prediction only	R	COG0313
PANA_3474	POORLY CHARACTERIZED	General function prediction only	R	COG3107
PANA_3475	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0792
PANA_3476	METABOLISM	Carbohydrate transport and metabolism	G	COG0279
PANA_3477	POORLY CHARACTERIZED	General function prediction only	R	COG2823
PANA_3478	POORLY CHARACTERIZED	General function prediction only	R	COG1741
PANA_3479	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3480	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3711
PANA_3481	NO COG	No COG association	NC	-
PANA_3482	METABOLISM	Amino acid transport and metabolism	E	COG1921
PANA_3483	POORLY CHARACTERIZED	General function prediction only	R	COG3964
PANA_3484	NO COG	No COG association	NC	-
PANA_3485	NO COG	No COG association	NC	-
PANA_3500	METABOLISM	Energy production and conversion	C	COG3783
PANA_3501	POORLY CHARACTERIZED	General function prediction only	R	COG0312
PANA_3502	POORLY CHARACTERIZED	Function unknown	S	COG3028
PANA_3503	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0773
PANA_3504	METABOLISM	Carbohydrate transport and metabolism	G	COG0158
PANA_3505	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_3505	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840
PANA_3507	POORLY CHARACTERIZED	Function unknown	S	COG2707
PANA_3508	METABOLISM	Energy production and conversion	C	COG0221
PANA_3510	POORLY CHARACTERIZED	Function unknown	S	COG2105
PANA_3511	POORLY CHARACTERIZED	Function unknown	S	COG2911
PANA_3512	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0729
PANA_3513	POORLY CHARACTERIZED	General function prediction only	R	COG1253
PANA_3514	NO COG	No COG association	NC	-
PANA_3515	POORLY CHARACTERIZED	General function prediction only	R	COG3054
PANA_3516	METABOLISM	Inorganic ion transport and metabolism	P	COG1218
PANA_3517	METABOLISM	Nucleotide transport and metabolism	F	COG0737
PANA_3518	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0545
PANA_3519	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3061
PANA_3521	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0359
PANA_3522	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0238
PANA_3523	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2965
PANA_3524	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0360
PANA_3525	POORLY CHARACTERIZED	General function prediction only	R	COG1073
PANA_3526	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_3526	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840
PANA_3527	NO COG	No COG association	NC	-
PANA_3528	METABOLISM	Lipid transport and metabolism	I	COG1960
PANA_3529	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0566
PANA_3530	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0557
PANA_3531	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1959
PANA_3532	METABOLISM	Nucleotide transport and metabolism	F	COG0104
PANA_3533	POORLY CHARACTERIZED	Function unknown	S	COG3242
PANA_3534	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0330
PANA_3535	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0330
PANA_3536	POORLY CHARACTERIZED	General function prediction only	R	COG2262
PANA_3537	POORLY CHARACTERIZED	General function prediction only	R	COG1923
PANA_3538	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0324

PANA_3539	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0323
PANA_3540	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0860
PANA_3541	POORLY CHARACTERIZED	General function prediction only	R	COG0802
PANA_3542	METABOLISM	Carbohydrate transport and metabolism	G	COG0063
PANA_3543	METABOLISM	Energy production and conversion	C	COG1600
PANA_3544	INFORMATION STORAGE AND PROCESSING	RNA processing and modification	A	COG1949
PANA_3545	POORLY CHARACTERIZED	General function prediction only	R	COG1162
PANA_3546	METABOLISM	Lipid transport and metabolism	I	COG0688
PANA_3547	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3264
PANA_3548	METABOLISM	Energy production and conversion	C	COG0039
PANA_3549	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1438
PANA_3550	NO COG	No COG association	NC	-
PANA_3556	POORLY CHARACTERIZED	Function unknown	S	COG1289
PANA_3557	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1566
PANA_3558	NO COG	No COG association	NC	-
PANA_3559	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3560	POORLY CHARACTERIZED	General function prediction only	R	COG0312
PANA_3561	POORLY CHARACTERIZED	Function unknown	S	COG3164
PANA_3562	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1530
PANA_3563	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG0424
PANA_3564	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2891
PANA_3565	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1792
PANA_3566	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG1077
PANA_3567	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2200
PANA_3568	POORLY CHARACTERIZED	General function prediction only	R	COG2041
PANA_3569	POORLY CHARACTERIZED	Function unknown	S	COG2717
PANA_3570	METABOLISM	Amino acid transport and metabolism	E	COG0757
PANA_3571	METABOLISM	Lipid transport and metabolism	I	COG0511
PANA_3572	METABOLISM	Lipid transport and metabolism	I	COG0439
PANA_3573	POORLY CHARACTERIZED	Function unknown	S	COG3924
PANA_3574	METABOLISM	Coenzyme transport and metabolism	H	COG4145
PANA_3575	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2264
PANA_3576	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0042
PANA_3577	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2901
PANA_3577	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2901
PANA_3578	METABOLISM	Amino acid transport and metabolism	E	COG0834
PANA_3578	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0834
PANA_3579	METABOLISM	Amino acid transport and metabolism	E	COG4597
PANA_3580	METABOLISM	Amino acid transport and metabolism	E	COG0765
PANA_3581	METABOLISM	Amino acid transport and metabolism	E	COG1126
PANA_3582	POORLY CHARACTERIZED	General function prediction only	R	COG0663
PANA_3584	METABOLISM	Amino acid transport and metabolism	E	COG0169
PANA_3585	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0009
PANA_3586	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0551
PANA_3587	POORLY CHARACTERIZED	Function unknown	S	COG2922
PANA_3588	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0758
PANA_3588	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0758
PANA_3589	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0242
PANA_3590	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0223
PANA_3591	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0144
PANA_3592	METABOLISM	Inorganic ion transport and metabolism	P	COG0569
PANA_3593	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1970
PANA_3595	NO COG	No COG association	NC	-
PANA_3596	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0203
PANA_3597	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0202
PANA_3598	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0522
PANA_3599	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0100
PANA_3600	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0099
PANA_3602	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0201
PANA_3603	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0200
PANA_3604	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1841
PANA_3605	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0098
PANA_3606	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0256
PANA_3607	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0097

PANA_3608	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0096
PANA_3609	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0199
PANA_3610	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0094
PANA_3611	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0198
PANA_3612	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0093
PANA_3613	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0186
PANA_3614	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0255
PANA_3615	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0197
PANA_3616	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0092
PANA_3617	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0091
PANA_3618	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0185
PANA_3619	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0090
PANA_3620	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0089
PANA_3621	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0088
PANA_3623	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0087
PANA_3624	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0051
PANA_3626	METABOLISM	Inorganic ion transport and metabolism	P	COG2193
PANA_3627	METABOLISM	Inorganic ion transport and metabolism	P	COG2906
PANA_3628	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0050
PANA_3629	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0480
PANA_3630	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0049
PANA_3631	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0048
PANA_3632	METABOLISM	Inorganic ion transport and metabolism	P	COG2168
PANA_3634	METABOLISM	Inorganic ion transport and metabolism	P	COG1553
PANA_3635	POORLY CHARACTERIZED	Function unknown	S	COG2964
PANA_3636	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0545
PANA_3637	POORLY CHARACTERIZED	Function unknown	S	COG2900
PANA_3638	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1047
PANA_3639	POORLY CHARACTERIZED	General function prediction only	R	COG3529
PANA_3640	METABOLISM	Inorganic ion transport and metabolism	P	COG0475
PANA_3641	POORLY CHARACTERIZED	General function prediction only	R	COG2249
PANA_3642	POORLY CHARACTERIZED	General function prediction only	R	COG0488
PANA_3643	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG2175
PANA_3644	METABOLISM	Inorganic ion transport and metabolism	P	COG0600
PANA_3645	METABOLISM	Inorganic ion transport and metabolism	P	COG4525
PANA_3646	METABOLISM	Inorganic ion transport and metabolism	P	COG4521
PANA_3647	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_3648	METABOLISM	Amino acid transport and metabolism	E	COG1280
PANA_3649	POORLY CHARACTERIZED	General function prediction only	R	COG0429
PANA_3650	POORLY CHARACTERIZED	Function unknown	S	COG3089
PANA_3651	METABOLISM	Energy production and conversion	C	COG3954
PANA_3652	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1765
PANA_3653	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0664
PANA_3654	POORLY CHARACTERIZED	Function unknown	S	COG1289
PANA_3655	METABOLISM	Amino acid transport and metabolism	E	COG4992
PANA_3656	METABOLISM	Amino acid transport and metabolism	E	COG0512
PANA_3656	METABOLISM	Coenzyme transport and metabolism	H	COG0512
PANA_3657	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG2184
PANA_3658	NO COG	No COG association	NC	-
PANA_3659	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0652
PANA_3660	METABOLISM	Carbohydrate transport and metabolism	G	COG0738
PANA_3661	METABOLISM	Nucleotide transport and metabolism	F	COG0402
PANA_3661	POORLY CHARACTERIZED	General function prediction only	R	COG0402
PANA_3662	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0180
PANA_3663	POORLY CHARACTERIZED	General function prediction only	R	COG0546
PANA_3664	METABOLISM	Carbohydrate transport and metabolism	G	COG0036
PANA_3665	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0338
PANA_3666	POORLY CHARACTERIZED	Function unknown	S	COG3266
PANA_3667	METABOLISM	Amino acid transport and metabolism	E	COG0337
PANA_3668	METABOLISM	Amino acid transport and metabolism	E	COG0703
PANA_3669	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG4796
PANA_3671	NO COG	No COG association	NC	-
PANA_3672	NO COG	No COG association	NC	-
PANA_3673	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG4972

PANA_3673	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG4972
PANA_3674	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG5009
PANA_3675	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0494
PANA_3675	POORLY CHARACTERIZED	General function prediction only	R	COG0494
PANA_3676	NO COG	No COG association	NC	-
PANA_3677	POORLY CHARACTERIZED	General function prediction only	R	COG1011
PANA_3678	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1188
PANA_3679	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1281
PANA_3680	METABOLISM	Energy production and conversion	C	COG1866
PANA_3681	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_3682	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_3682	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_3683	NO COG	No COG association	NC	-
PANA_3684	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0782
PANA_3685	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2183
PANA_3686	METABOLISM	Carbohydrate transport and metabolism	G	COG0366
PANA_3687	NO COG	No COG association	NC	-
PANA_3688	POORLY CHARACTERIZED	General function prediction only	R	COG0596
PANA_3689	POORLY CHARACTERIZED	General function prediction only	R	COG1040
PANA_3690	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0694
PANA_3691	METABOLISM	Carbohydrate transport and metabolism	G	COG1640
PANA_3692	METABOLISM	Carbohydrate transport and metabolism	G	COG0058
PANA_3693	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1349
PANA_3693	METABOLISM	Carbohydrate transport and metabolism	G	COG1349
PANA_3694	POORLY CHARACTERIZED	General function prediction only	R	COG0705
PANA_3695	METABOLISM	Inorganic ion transport and metabolism	P	COG0607
PANA_3696	METABOLISM	Energy production and conversion	C	COG0578
PANA_3697	METABOLISM	Carbohydrate transport and metabolism	G	COG0058
PANA_3698	METABOLISM	Carbohydrate transport and metabolism	G	COG0297
PANA_3699	METABOLISM	Carbohydrate transport and metabolism	G	COG0448
PANA_3700	METABOLISM	Carbohydrate transport and metabolism	G	COG1523
PANA_3701	METABOLISM	Carbohydrate transport and metabolism	G	COG0296
PANA_3702	METABOLISM	Amino acid transport and metabolism	E	COG0136
PANA_3703	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG2095
PANA_3704	METABOLISM	Carbohydrate transport and metabolism	G	COG2610
PANA_3704	METABOLISM	Amino acid transport and metabolism	E	COG2610
PANA_3705	METABOLISM	Carbohydrate transport and metabolism	G	COG3265
PANA_3706	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_3707	POORLY CHARACTERIZED	General function prediction only	R	COG1741
PANA_3708	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1247
PANA_3709	METABOLISM	Amino acid transport and metabolism	E	COG0405
PANA_3710	NO COG	No COG association	NC	-
PANA_3711	METABOLISM	Energy production and conversion	C	COG0584
PANA_3712	METABOLISM	Carbohydrate transport and metabolism	G	COG3839
PANA_3713	METABOLISM	Carbohydrate transport and metabolism	G	COG0395
PANA_3714	METABOLISM	Carbohydrate transport and metabolism	G	COG1175
PANA_3715	METABOLISM	Carbohydrate transport and metabolism	G	COG1653
PANA_3716	METABOLISM	Amino acid transport and metabolism	E	COG0410
PANA_3717	METABOLISM	Amino acid transport and metabolism	E	COG0411
PANA_3718	METABOLISM	Amino acid transport and metabolism	E	COG4177
PANA_3719	METABOLISM	Amino acid transport and metabolism	E	COG0559
PANA_3720	METABOLISM	Amino acid transport and metabolism	E	COG0683
PANA_3721	NO COG	No COG association	NC	-
PANA_3722	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0568
PANA_3723	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG2177
PANA_3724	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG2884
PANA_3725	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG0552
PANA_3726	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0742
PANA_3727	POORLY CHARACTERIZED	Function unknown	S	COG3776
PANA_3728	NO COG	No COG association	NC	-
PANA_3729	POORLY CHARACTERIZED	Function unknown	S	COG3714
PANA_3730	METABOLISM	Inorganic ion transport and metabolism	P	COG2217
PANA_3731	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0425
PANA_3732	POORLY CHARACTERIZED	Function unknown	S	COG1738

PANA_3733	NO COG	No COG association	NC	-
PANA_3734	POORLY CHARACTERIZED	Function unknown	S	COG3892
PANA_3735	METABOLISM	Amino acid transport and metabolism	E	COG3962
PANA_3736	POORLY CHARACTERIZED	General function prediction only	R	COG0673
PANA_3737	METABOLISM	Carbohydrate transport and metabolism	G	COG1082
PANA_3738	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1737
PANA_3740	METABOLISM	Carbohydrate transport and metabolism	G	COG3718
PANA_3741	METABOLISM	Energy production and conversion	C	COG1012
PANA_3742	METABOLISM	Nucleotide transport and metabolism	F	COG2233
PANA_3746	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2200
PANA_3748	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_3749	METABOLISM	Coenzyme transport and metabolism	H	COG0113
PANA_3750	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG2207
PANA_3751	POORLY CHARACTERIZED	Function unknown	S	COG1636
PANA_3756	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG2173
PANA_3760	METABOLISM	Amino acid transport and metabolism	E	COG0444
PANA_3760	METABOLISM	Inorganic ion transport and metabolism	P	COG0444
PANA_3763	POORLY CHARACTERIZED	General function prediction only	R	COG2081
PANA_3764	METABOLISM	Inorganic ion transport and metabolism	P	COG0306
PANA_3765	NO COG	No COG association	NC	-
PANA_3766	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0589
PANA_3767	NO COG	No COG association	NC	-
PANA_3768	METABOLISM	Amino acid transport and metabolism	E	COG0339
PANA_3769	POORLY CHARACTERIZED	General function prediction only	R	COG2961
PANA_3771	POORLY CHARACTERIZED	General function prediction only	R	COG1234
PANA_3772	METABOLISM	Energy production and conversion	C	COG1249
PANA_3781	METABOLISM	Lipid transport and metabolism	I	COG0657
PANA_3784	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_3785	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3786	POORLY CHARACTERIZED	Function unknown	S	COG2862
PANA_3787	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2205
PANA_3788	METABOLISM	Inorganic ion transport and metabolism	P	COG2156
PANA_3789	METABOLISM	Inorganic ion transport and metabolism	P	COG2216
PANA_3790	METABOLISM	Inorganic ion transport and metabolism	P	COG2060
PANA_3792	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2206
PANA_3793	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1670
PANA_3794	CELLULAR PROCESSING AND SIGNALLING	Defence mechanisms	V	COG1132
PANA_3795	POORLY CHARACTERIZED	Function unknown	S	COG3146
PANA_3796	POORLY CHARACTERIZED	General function prediction only	R	COG0400
PANA_3798	METABOLISM	Nucleotide transport and metabolism	F	COG1957
PANA_3799	METABOLISM	Amino acid transport and metabolism	E	COG1063
PANA_3799	POORLY CHARACTERIZED	General function prediction only	R	COG1063
PANA_3800	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1802
PANA_3801	METABOLISM	Carbohydrate transport and metabolism	G	COG2271
PANA_3805	POORLY CHARACTERIZED	Function unknown	S	COG1971
PANA_3807	POORLY CHARACTERIZED	General function prediction only	R	COG0579
PANA_3808	METABOLISM	Amino acid transport and metabolism	E	COG1280
PANA_3809	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1167
PANA_3809	METABOLISM	Amino acid transport and metabolism	E	COG1167
PANA_3810	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0625
PANA_3811	POORLY CHARACTERIZED	General function prediction only	R	COG0384
PANA_3812	METABOLISM	Lipid transport and metabolism	I	COG1028
PANA_3812	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1028
PANA_3812	POORLY CHARACTERIZED	General function prediction only	R	COG1028
PANA_3813	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3814	POORLY CHARACTERIZED	Function unknown	S	COG1295
PANA_3815	METABOLISM	Carbohydrate transport and metabolism	G	COG2211
PANA_3823	METABOLISM	Amino acid transport and metabolism	E	COG0059
PANA_3823	METABOLISM	Coenzyme transport and metabolism	H	COG0059
PANA_3824	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3825	METABOLISM	Amino acid transport and metabolism	E	COG1171
PANA_3826	METABOLISM	Amino acid transport and metabolism	E	COG0129
PANA_3826	METABOLISM	Carbohydrate transport and metabolism	G	COG0129
PANA_3827	METABOLISM	Amino acid transport and metabolism	E	COG0115

PANA_3827	METABOLISM	Coenzyme transport and metabolism	H	COG0115
PANA_3828	POORLY CHARACTERIZED	Function unknown	S	COG3978
PANA_3829	METABOLISM	Amino acid transport and metabolism	E	COG0028
PANA_3829	METABOLISM	Coenzyme transport and metabolism	H	COG0028
PANA_3830	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0606
PANA_3831	POORLY CHARACTERIZED	Function unknown	S	COG3085
PANA_3832	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3833	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0796
PANA_3834	METABOLISM	Coenzyme transport and metabolism	H	COG4206
PANA_3835	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG2265
PANA_3836	NO COG	No COG association	NC	-
PANA_3837	NO COG	No COG association	NC	-
PANA_3838	METABOLISM	Energy production and conversion	C	COG1249
PANA_3839	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3840	METABOLISM	Amino acid transport and metabolism	E	COG0165
PANA_3841	METABOLISM	Amino acid transport and metabolism	E	COG0137
PANA_3842	METABOLISM	Amino acid transport and metabolism	E	COG0548
PANA_3843	METABOLISM	Amino acid transport and metabolism	E	COG0002
PANA_3844	METABOLISM	Amino acid transport and metabolism	E	COG0624
PANA_3845	METABOLISM	Energy production and conversion	C	COG2352
PANA_3847	METABOLISM	Amino acid transport and metabolism	E	COG0685
PANA_3848	METABOLISM	Amino acid transport and metabolism	E	COG0527
PANA_3849	METABOLISM	Amino acid transport and metabolism	E	COG0626
PANA_3850	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG3060
PANA_3850	METABOLISM	Amino acid transport and metabolism	E	COG3060
PANA_3851	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0254
PANA_3852	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1198
PANA_3853	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_3854	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG3087
PANA_3855	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG5405
PANA_3856	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1220
PANA_3857	METABOLISM	Coenzyme transport and metabolism	H	COG0684
PANA_3858	POORLY CHARACTERIZED	Function unknown	S	COG3074
PANA_3859	METABOLISM	Carbohydrate transport and metabolism	G	COG0580
PANA_3860	METABOLISM	Energy production and conversion	C	COG0554
PANA_3861	METABOLISM	Carbohydrate transport and metabolism	G	COG1494
PANA_3862	METABOLISM	Carbohydrate transport and metabolism	G	COG2814
PANA_3863	METABOLISM	Energy production and conversion	C	COG1018
PANA_3865	NO COG	No COG association	NC	-
PANA_3866	METABOLISM	Carbohydrate transport and metabolism	G	COG0149
PANA_3867	METABOLISM	Lipid transport and metabolism	I	COG2134
PANA_3868	METABOLISM	Inorganic ion transport and metabolism	P	COG1613
PANA_3869	METABOLISM	Carbohydrate transport and metabolism	G	COG0205
PANA_3870	METABOLISM	Inorganic ion transport and metabolism	P	COG0053
PANA_3873	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG3678
PANA_3873	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG3678
PANA_3873	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3678
PANA_3873	METABOLISM	Inorganic ion transport and metabolism	P	COG3678
PANA_3874	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_3874	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_3875	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_3876	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0219
PANA_3877	METABOLISM	Amino acid transport and metabolism	E	COG1045
PANA_3878	METABOLISM	Energy production and conversion	C	COG0240
PANA_3879	CELLULAR PROCESSING AND SIGNALLING	Intracellular trafficking, secretion and vesicular transport	U	COG1952
PANA_3880	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG0695
PANA_3881	METABOLISM	Inorganic ion transport and metabolism	P	COG0607
PANA_3882	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG4942
PANA_3883	POORLY CHARACTERIZED	Function unknown	S	COG2861
PANA_3884	METABOLISM	Amino acid transport and metabolism	E	COG1063
PANA_3884	POORLY CHARACTERIZED	General function prediction only	R	COG1063
PANA_3885	METABOLISM	Coenzyme transport and metabolism	H	COG0156
PANA_3886	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0451
PANA_3886	METABOLISM	Carbohydrate transport and metabolism	G	COG0451

PANA_3887	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0859
PANA_3888	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0859
PANA_3895	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0463
PANA_3896	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0859
PANA_3897	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0438
PANA_3898	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1519
PANA_3899	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0463
PANA_3900	METABOLISM	Coenzyme transport and metabolism	H	COG0669
PANA_3901	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0266
PANA_3902	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0267
PANA_3903	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0227
PANA_3904	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG2003
PANA_3905	METABOLISM	Coenzyme transport and metabolism	H	COG0452
PANA_3906	METABOLISM	Nucleotide transport and metabolism	F	COG0756
PANA_3907	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1309
PANA_3908	METABOLISM	Nucleotide transport and metabolism	F	COG0461
PANA_3909	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0689
PANA_3910	METABOLISM	Nucleotide transport and metabolism	F	COG1972
PANA_3911	NO COG	No COG association	NC	-
PANA_3912	POORLY CHARACTERIZED	Function unknown	S	COG1561
PANA_3932	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1560
PANA_3933	POORLY CHARACTERIZED	General function prediction only	R	COG2244
PANA_3934	POORLY CHARACTERIZED	Function unknown	S	COG2860
PANA_3935	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0272
PANA_3936	METABOLISM	Nucleotide transport and metabolism	F	COG0194
PANA_3937	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1758
PANA_3938	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0317
PANA_3938	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0317
PANA_3939	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0566
PANA_3940	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG1200
PANA_3940	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1200
PANA_3941	METABOLISM	Nucleotide transport and metabolism	F	COG2233
PANA_3942	NO COG	No COG association	NC	-
PANA_3943	POORLY CHARACTERIZED	General function prediction only	R	COG2153
PANA_3944	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1490
PANA_3945	POORLY CHARACTERIZED	Function unknown	S	COG1295
PANA_3946	POORLY CHARACTERIZED	General function prediction only	R	COG1011
PANA_3947	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG1217
PANA_3949	METABOLISM	Amino acid transport and metabolism	E	COG0174
PANA_3950	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3852
PANA_3951	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2204
PANA_3952	METABOLISM	Coenzyme transport and metabolism	H	COG0635
PANA_3953	POORLY CHARACTERIZED	Function unknown	S	COG3078
PANA_3954	POORLY CHARACTERIZED	General function prediction only	R	COG0218
PANA_3955	INFORMATION STORAGE AND PROCESSING	Replication, recombination and repair	L	COG0749
PANA_3956	NO COG	No COG association	NC	-
PANA_3957	POORLY CHARACTERIZED	General function prediction only	R	COG2334
PANA_3958	POORLY CHARACTERIZED	Function unknown	S	COG3084
PANA_3959	METABOLISM	Coenzyme transport and metabolism	H	COG0746
PANA_3961	METABOLISM	Nucleotide transport and metabolism	F	COG0127
PANA_3962	NO COG	No COG association	NC	-
PANA_3963	POORLY CHARACTERIZED	Function unknown	S	COG3141
PANA_3964	METABOLISM	Carbohydrate transport and metabolism	G	COG0662
PANA_3965	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_3965	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840
PANA_3966	METABOLISM	Amino acid transport and metabolism	E	COG0334
PANA_3967	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_3968	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG3527
PANA_3969	METABOLISM	Amino acid transport and metabolism	E	COG0028
PANA_3969	METABOLISM	Coenzyme transport and metabolism	H	COG0028
PANA_3972	POORLY CHARACTERIZED	Function unknown	S	COG3558
PANA_3973	METABOLISM	Coenzyme transport and metabolism	H	COG0054
PANA_3975	METABOLISM	Inorganic ion transport and metabolism	P	COG1464
PANA_3977	METABOLISM	Lipid transport and metabolism	I	COG1028

PANA_3977	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1028
PANA_3977	POORLY CHARACTERIZED	General function prediction only	R	COG1028
PANA_3979	METABOLISM	Carbohydrate transport and metabolism	G	COG4692
PANA_3980	METABOLISM	Amino acid transport and metabolism	E	COG0591
PANA_3980	POORLY CHARACTERIZED	General function prediction only	R	COG0591
PANA_3982	METABOLISM	Energy production and conversion	C	COG1454
PANA_3983	METABOLISM	Coenzyme transport and metabolism	H	COG1995
PANA_3990	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG4977
PANA_3991	METABOLISM	Carbohydrate transport and metabolism	G	COG0697
PANA_3991	METABOLISM	Amino acid transport and metabolism	E	COG0697
PANA_3991	POORLY CHARACTERIZED	General function prediction only	R	COG0697
PANA_3992	POORLY CHARACTERIZED	Function unknown	S	COG2105
PANA_3993	CELLULAR PROCESSING AND SIGNALLING	Cell motility	N	COG0840
PANA_3993	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0840
PANA_3995	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0845
PANA_3996	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0642
PANA_3997	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG0745
PANA_3997	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0745
PANA_3998	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG3713
PANA_4002	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0399
PANA_4003	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0463
PANA_4004	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0223
PANA_4005	METABOLISM	Carbohydrate transport and metabolism	G	COG0726
PANA_4006	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1807
PANA_4007	METABOLISM	Inorganic ion transport and metabolism	P	COG2076
PANA_4008	NO COG	No COG association	NC	-
PANA_4022	NO COG	No COG association	NC	-
PANA_4028	POORLY CHARACTERIZED	Function unknown	S	COG5654
PANA_4029	POORLY CHARACTERIZED	Function unknown	S	COG5642
PANA_4033	METABOLISM	Energy production and conversion	C	COG2141
PANA_4034	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1335
PANA_4035	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0251
PANA_4038	POORLY CHARACTERIZED	General function prediction only	R	COG1853
PANA_4039	METABOLISM	Nucleotide transport and metabolism	F	COG2233
PANA_4040	POORLY CHARACTERIZED	Function unknown	S	COG3367
PANA_4041	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG4948
PANA_4041	POORLY CHARACTERIZED	General function prediction only	R	COG4948
PANA_4042	POORLY CHARACTERIZED	General function prediction only	R	COG1123
PANA_4043	METABOLISM	Amino acid transport and metabolism	E	COG0747
PANA_4044	METABOLISM	Amino acid transport and metabolism	E	COG0601
PANA_4044	METABOLISM	Inorganic ion transport and metabolism	P	COG0601
PANA_4045	METABOLISM	Amino acid transport and metabolism	E	COG1173
PANA_4045	METABOLISM	Inorganic ion transport and metabolism	P	COG1173
PANA_4046	METABOLISM	Amino acid transport and metabolism	E	COG3191
PANA_4046	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG3191
PANA_4047	METABOLISM	Amino acid transport and metabolism	E	COG2362
PANA_4048	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_4050	METABOLISM	Carbohydrate transport and metabolism	G	COG1879
PANA_4052	POORLY CHARACTERIZED	General function prediction only	R	COG0673
PANA_4053	METABOLISM	Carbohydrate transport and metabolism	G	COG1082
PANA_4072	METABOLISM	Lipid transport and metabolism	I	COG4770
PANA_4073	POORLY CHARACTERIZED	Function unknown	S	COG3665
PANA_4074	POORLY CHARACTERIZED	Function unknown	S	COG3665
PANA_4075	METABOLISM	Inorganic ion transport and metabolism	P	COG1116
PANA_4076	METABOLISM	Inorganic ion transport and metabolism	P	COG0600
PANA_4077	NO COG	No COG association	NC	-
PANA_4078	POORLY CHARACTERIZED	Function unknown	S	COG4291
PANA_4079	NO COG	No COG association	NC	-
PANA_4083	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG1794
PANA_4084	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG4232
PANA_4084	METABOLISM	Energy production and conversion	C	COG4232
PANA_4085	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1225
PANA_4087	NO COG	No COG association	NC	-
PANA_4089	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583

PANA_4105	METABOLISM	Coenzyme transport and metabolism	H	COG0095
PANA_4108	METABOLISM	Amino acid transport and metabolism	E	COG0160
PANA_4109	METABOLISM	Amino acid transport and metabolism	E	COG0076
PANA_4110	NO COG	No COG association	NC	-
PANA_4112	POORLY CHARACTERIZED	General function prediction only	R	COG0579
PANA_4113	POORLY CHARACTERIZED	Function unknown	S	COG5464
PANA_4114	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG3290
PANA_4115	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG4565
PANA_4115	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG4565
PANA_4116	METABOLISM	Coenzyme transport and metabolism	H	COG1477
PANA_4117	METABOLISM	Energy production and conversion	C	COG1902
PANA_4118	METABOLISM	Inorganic ion transport and metabolism	P	COG0471
PANA_4119	METABOLISM	Energy production and conversion	C	COG1951
PANA_4121	CELLULAR PROCESSING AND SIGNALLING	Cell wall/membrane/envelope biogenesis	M	COG0451
PANA_4121	METABOLISM	Carbohydrate transport and metabolism	G	COG0451
PANA_4122	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG0583
PANA_4128	POORLY CHARACTERIZED	General function prediction only	R	COG1881
PANA_4153	METABOLISM	Amino acid transport and metabolism	E	COG2066
PANA_4154	METABOLISM	Energy production and conversion	C	COG1052
PANA_4154	METABOLISM	Coenzyme transport and metabolism	H	COG1052
PANA_4154	POORLY CHARACTERIZED	General function prediction only	R	COG1052
PANA_4158	METABOLISM	Coenzyme transport and metabolism	H	COG0142
PANA_4160	METABOLISM	Coenzyme transport and metabolism	H	COG0654
PANA_4160	METABOLISM	Energy production and conversion	C	COG0654
PANA_4161	METABOLISM	Secondary metabolites biosynthesis, transport and catabolism	Q	COG1233
PANA_4162	METABOLISM	Lipid transport and metabolism	I	COG1562
PANA_4163	NO COG	No COG association	NC	-
PANA_4164	POORLY CHARACTERIZED	General function prediction only	R	COG1064
PANA_4166	NO COG	No COG association	NC	-
PANA_4171	METABOLISM	Amino acid transport and metabolism	E	COG1280
PANA_4172	NO COG	No COG association	NC	-
PANA_4173	METABOLISM	Coenzyme transport and metabolism	H	COG0476
PANA_4174	METABOLISM	Coenzyme transport and metabolism	H	COG2022
PANA_4175	METABOLISM	Coenzyme transport and metabolism	H	COG2104
PANA_4176	METABOLISM	Amino acid transport and metabolism	E	COG0665
PANA_4177	NO COG	No COG association	NC	-
PANA_4178	METABOLISM	Amino acid transport and metabolism	E	COG1296
PANA_4179	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1396
PANA_4184	POORLY CHARACTERIZED	General function prediction only	R	COG0523
PANA_4185	METABOLISM	Energy production and conversion	C	COG1012
PANA_4186	NO COG	No COG association	NC	-
PANA_4187	CELLULAR PROCESSING AND SIGNALLING	Cell cycle control, cell division, chromosome partitioning	D	COG1192
PANA_4188	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1475
PANA_4193	NO COG	No COG association	NC	-
PANA_4195	NO COG	No COG association	NC	-
PANA_4196	NO COG	No COG association	NC	-
PANA_4197	NO COG	No COG association	NC	-
PANA_4198	NO COG	No COG association	NC	-
PANA_4199	NO COG	No COG association	NC	-
PANA_4200	METABOLISM	Inorganic ion transport and metabolism	P	COG0803
PANA_4201	METABOLISM	Inorganic ion transport and metabolism	P	COG1121
PANA_4202	METABOLISM	Inorganic ion transport and metabolism	P	COG1108
PANA_4203	METABOLISM	Inorganic ion transport and metabolism	P	COG1108
PANA_4204	METABOLISM	Energy production and conversion	C	COG2080
PANA_4208	CELLULAR PROCESSING AND SIGNALLING	Post-translational modification, protein turnover and chaperones	O	COG1975
PANA_4209	METABOLISM	Inorganic ion transport and metabolism	P	COG0614
PANA_4210	METABOLISM	Inorganic ion transport and metabolism	P	COG4773
PANA_4211	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1737
PANA_4213	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG1187
PANA_4217	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2200
PANA_4218	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1349
PANA_4218	METABOLISM	Carbohydrate transport and metabolism	G	COG1349
PANA_4222	METABOLISM	Carbohydrate transport and metabolism	G	COG2723
PANA_4223	METABOLISM	Carbohydrate transport and metabolism	G	COG1263

PANA_4224	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1609
PANA_4225	INFORMATION STORAGE AND PROCESSING	Transcription	K	COG1349
PANA_4225	METABOLISM	Carbohydrate transport and metabolism	G	COG1349
PANA_4226	NO COG	No COG association	NC	-
PANA_4227	POORLY CHARACTERIZED	General function prediction only	R	COG0561
PANA_4228	NO COG	No COG association	NC	-
PANA_4229	NO COG	No COG association	NC	-
PANA_4230	CELLULAR PROCESSING AND SIGNALLING	Signal transduction mechanisms	T	COG2199
PANA_4231	INFORMATION STORAGE AND PROCESSING	Translation, ribosomal structure and biogenesis	J	COG0024
PANA_4232	NO COG	No COG association	NC	-
PANA_4237	METABOLISM	Inorganic ion transport and metabolism	P	COG4773

Appendix B

Table 1 Shimodaira-Hasegawa test results for the cellular processing and signalling (Cellular) dataset

Tree Topology	Likelihood value	D(Likelihood)	SD	Significantly worse *
Cellular	- 475579.542	-	-	Obtained
Core	- 475399.338	180.204	84.231	Best
Information	- 475579.542	0.000	0.002	No
Metabolism	- 475399.338	180.204	84.231	No
No COG	- 475610.538	- 30.996	54.067	No
Poorly characterised	- 475795.589	- 216.047	72.246	Yes
Ribosomal	- 475967.087	- 387.545	63.047	Yes
Inferred MLSA	- 475776.897	- 197.355	41.537	Yes

*Significance level of 1% (p-value = 0.01)

Table 2 Shimodaira-Hasegawa test results for the dataset of the full core genome (Core)

*Significance level of 1% (p-value = 0.01)

Tree Topology	Likelihood value	D(Likelihood)	SD	Significantly worse *
Cellular	- 2161973.197	- 772.874	146.164	Yes
Core	- 2161200.322	-	-	Obtained/Best
Information	- 2161973.197	- 772.874	146.164	Yes
Metabolism	- 2161200.322	0.000	0.004	No
No COG	- 2161781.268	- 580.946	93.170	Yes
Poorly characterised	- 2161411.781	- 211.459	101.133	No
Ribosomal	- 2162582.617	- 1382.295	173.708	Yes
Inferred MLSA	- 2162936.915	- 1736.592	169.643	Yes

Table 3 Shimodaira-Hasegawa test results for the information storage and processing (Information) dataset

Tree Topology	Likelihood value	D(Likelihood)	SD	Significantly worse *
Cellular	- 445971.106	0.000	0.001	No
Core	- 445847.266	123.840	62.961	Best
Information	- 445971.106	-	-	Obtained
Metabolism	- 445847.266	123.840	62.961	No
No COG	- 445983.990	- 12.885	53.129	No
Poorly characterised	- 445854.846	116.260	62.741	No
Ribosomal	- 446061.775	- 90.669	53.295	No
Inferred MLSA	- 446191.447	- 220.342	41.778	Yes

*Significance level of 1% (p-value = 0.01)

Table 4 Shimodaira-Hasegawa results for the metabolism dataset

Tree Topology	Likelihood value	D(Likelihood)	SD	Significantly worse *
Cellular	- 960919.004	- 497.972	89.498	Yes
Core	- 960421.032	0.000	0.004	No
Information	- 960919.004	- 497.972	89.498	Yes
Metabolism	- 960421.032	-	-	Obtained
No COG	- 960719.340	- 298.307	52.322	Yes
Poorly characterised	- 960351.268	69.764	65.833	Best
Ribosomal	- 960973.073	- 552.040	112.379	Yes
Inferred MLSA	- 961330.325	- 909.293	105.369	Yes

*Significance level of 1% (p-value = 0.01)

Table 5 Shimodaira-Hasegawa results for the dataset for genes with no COG database homologues (No COG)

Tree Topology	Likelihood value	D(Likelihood)	SD	Significantly worse *
Cellular	- 81974.127	- 5.437	22.117	No
Core	- 82078.602	- 109.912	24.245	Yes
Information	- 81974.127	- 5.437	22.117	No
Metabolism	- 82078.602	- 109.912	24.245	Yes
No COG	- 81968.689	-	-	Obtained/Best
Poorly characterised	- 82052.046	- 83.356	26.577	Yes
Ribosomal	- 82123.384	- 154.695	37.961	Yes
Inferred MLSA	- 82038.473	- 69.783	27.557	No

*Significance level of 1% (p-value = 0.01)

Table 6 Shimodaira-Hasegawa results for the poorly characterised dataset

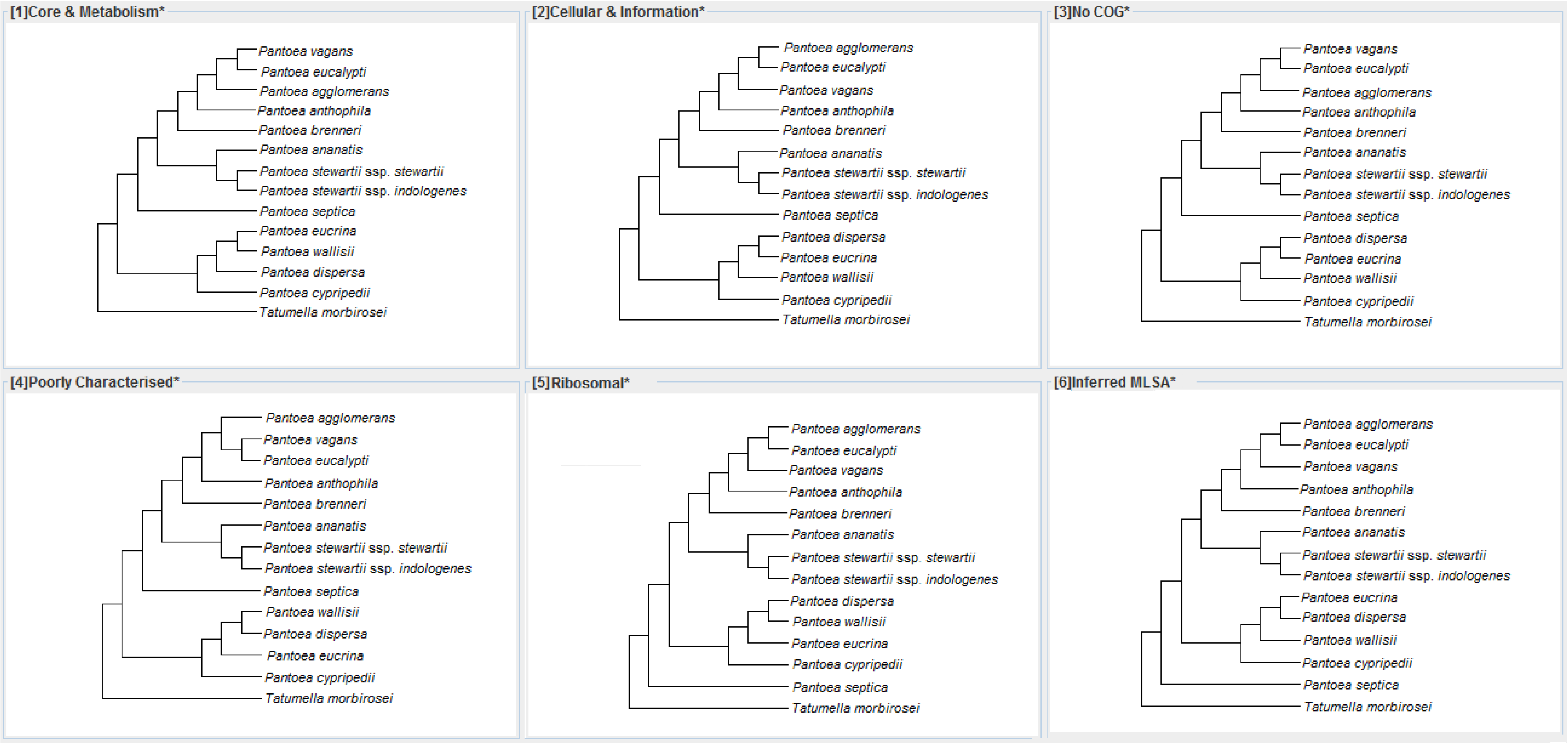
Tree Topology	Likelihood value	D(Likelihood)	SD	Significantly worse*
Cellular	- 264807.987	- 166.842	51.938	Yes
Core	- 264760.366	-119.222	34.737	Yes
Information	- 264807.987	- 166.842	51.938	Yes
Metabolism	- 264760.366	-119.222	34.737	Yes
No COG	- 264781.970	- 140.825	33.362	Yes
Poorly characterised	- 264641.145	-	-	Obtained/Best
Ribosomal	- 264768.312	-127.167	48.555	Yes
Inferred MLSA	- 264911.795	- 270.650	59.119	Yes

*Significance level of 1% (p-value = 0.01)

Table 7 Shimodaira-Hasegawa results for the structural ribosomal protein (Ribosomal) dataset

Tree Topology	Likelihood value	D(Likelihood)	SD	Significantly worse *
Cellular	- 25630.912	- 51.329	22.525	No
Core	- 25644.062	- 64.479	28.647	No
Information	- 25630.912	- 51.329	22.525	No
Metabolism	- 25644.062	- 64.479	28.647	No
No COG	- 25656.393	- 76.810	27.648	Yes
Poorly characterised	- 25613.130	- 33.546	22.536	No
Ribosomal	- 25579.584	-	-	Obtained/Best
Inferred MLSA	- 25628.671	- 49.088	16.521	Yes

*Significance level of 1% (p-value = 0.01)


[1]Core & Metabolism*

[2]Cellular & Information*

[3]No COG*

[4]Poorly Characterised*

[5]Ribosomal*

[6]Inferred MLSA*

Figure 1 Tree topologies used for the Shimodaira-Hasegawa tests with the different concatenated datasets. Trees were inferred from phylogenetic trees obtained in Chapter 3. Only a single isolate for each species were included in the inferred trees as the intraspecies relationships were not the focus of the study.