



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Natural and Agricultural Sciences

Comparative genomics study of completely sequenced *Thermus* sp. strains to enhance and facilitate their application in biotechnology

by

Benjamin Kumwenda

Submitted in partial fulfilment of the degree Philosophiae Doctor Bioinformatics

In the Faculty of Natural and Agricultural Sciences

Bioinformatics and Computational Biology Unit

Department of Biochemistry

University of Pretoria

November, 2013

Submission Declaration

I **Benjamin Kumwenda** declare that this thesis/dissertation, which I hereby submit for the degree of Philosophiae Doctor at the University of Pretoria is my own work and has not been previously submitted by me or anyone else for a degree at this or any other institution.

Benjamin Kumwenda

Full Name



Signature

30 November 2013

Date

Publications

Kumwenda, B., Litthauer, D., Tastan Bishop, O. & Reva, O. (2013) **Analysis of protein enhancing factors in industrial important *Thermus* bacteria.** *Evolutionary Bioinformatics*, 9:327-342.

Gounder, K., Brzuszkiewicz, E., Liesegang, H., Wollherr, A., Daniel, R., Gottschalk, G., Reva, O., Kumwenda, B., Srivastava, M., van Heerden, E. & Litthauer, D. (2011). **Sequence of the hyperplastic genome of the naturally competent *Thermus scotoductus* SA-01.** *BMC Genomics*, 12, 577.

Dedication

This PhD work is dedicated to they who laboured but did not live to see it; to my mother and father, with fond memories.

Acknowledgements

I would like to thank my supervisors, Prof. Oleg Reva and Prof. Derek Litthauer for guiding and encouraging me through this work. I am grateful to Dr Özlem Tastan Bishop of Rhodes University Bioinformatics who taught me protein structural modelling for the analysis in this work.

I am so thankful to my beautiful and lovely wife Grace Kumwenda for being faithful and untiringly supportive to me throughout my studies, without her support and encouragement this project could not have been possible. Dr and Mrs Kumwenda, thank you for inspiring and supporting me since my youth, through all my studies and to finally attain this Ph.D degree.

I am appreciative to Prof. Fourie Joubert, my fellow students, Oliver Bezuidt, Nannette Cotzer and the entire staff of University of Pretoria Bioinformatics and Computational Biology unit for providing me with a conducive and supportive environment for my work.

This work was possible by funding from Southern Africa Biotechnology and Informatics for Natural products and the National Research Foundation Grant 71261. The University of Malawi, College of Medicine was kind to grant me a paid four years study leave in support of my studies.

It is the almighty God, through Jesus Christ who made this possible. He is the one who gave me the life, health and strength to do this work; to Him be all the glory!

Abstract

Thermus bacteria are of special interest because of their ability to live in high temperature environments. Their enzymes exhibit higher and stable activity in industry as compared to mesophilic or synthetic counterparts. *Thermus* bacteria are capable of reducing heavy metals which is essential in eradication of heavy metal pollution and controlling global warming.

Genome rearrangements were investigated in *Thermus* species and the extent to which they affected the distribution of functionally related genes on the chromosome and its implication on the coherence of the metabolic network. The contribution of horizontal gene transfer to genome rearrangements and the shuffling of genes on the chromosome were analysed. Horizontally transferred genes were identified alongside their donors and recipients, their age and relative time of insertion. Metabolic networks were clustered and compared to determine the extent to which they were affected by rearrangements as a measure of evolutionary pressures experienced by organisms. Factors that enhance protein thermostability were analysed by determining dominant substitutions for amino acid residues and their properties. Protein thermostability was measured using the UNAFold algorithm. Amino acid substitutions were compared between less and highly thermophilic orthologous sequences in *T. scotoductus* SA-01 and *T. thermophilus* HB27 respectively. Protein structures were modelled for orthologs that met a defined selection criterion. Dominant amino acid substitutions were analysed in the structures to determine their locations. The contribution of dominant substitutions to energy changes in structures was analysed using FoldX program.

Results revealed a uniform distribution of functionally related genes among thermophilic and mesophilic organisms. The contribution of horizontal gene transfer to genome rearrangements was found to be insignificant. Metabolic networks for *Thermus* species were poorly clustered in correlation with their optimum environmental growth temperatures. Non-polar, small and charged amino acids were found to significantly enhance thermostability. Higher occurrence of alanine substituted by serine and threonine; and arginine substituted by glutamine and lysine were observed to influence thermostability. Structural comparison showed that mutations were mostly located on the surfaces and helices of structures. The positions of mutations appeared to influence their energy contribution to the overall structure as measured by FoldX algorithm.

Table of Contents

Submission Declaration.....	ii
Publications	iii
Dedication	iv
Acknowledgements	v
Abstract	vi
Table of Figures	x
Table of Tables.....	xi
List of Acronyms.....	xii
Introduction	1
1.1 Introduction	1
1.2 Research Problem.....	3
1.3 Aims and Objectives	4
1.4 Methodology	5
1.5 <i>Thermus</i> Bacteria.....	7
1.6 Industrial Application.....	9
1.6.1 Heavy metal reduction.....	10
1.6.2 Thermostable enzymes	13
1.7 Protein and Enzyme Thermostability	19
1.7.1 Factors enhancing protein thermostability	19
1.7.2 Approaches for measuring protein thermostability	26
1.8 Genome Rearrangements	31
1.8.1 Horizontal Gene Transfer.....	33
1.8.2 Breakpoints analysis.....	40
1.9 Metabolic networks	40
1.9.1 Construction of metabolic networks and databases.....	43
1.10 Summary and aims of the project.....	48
Genome Rearrangements	51
2.1 Introduction	51
2.2 Research Questions and Hypotheses	53
2.3 Analysis of the Distribution of Genes	55
2.3.1 Gene distribution data set	55
2.3.2 Methods.....	56
2.4 Breakpoints Analysis.....	59
2.4.1 Breakpoints dataset.....	60
2.4.2 Methods.....	61

2.5	Metabolic Network Clustering	64
2.5.1	Data for metabolic network clustering	67
2.6	Gene Distribution Results	67
2.7	Breakpoints Results	69
2.8	Metabolic Network Clustering Results	76
2.8.1	Clustering based on operons and genomic distance	76
2.9	Discussion	79
2.10	Summary	81
Horizontal Gene Transfer		83
3.1	Introduction	83
3.2	Aim.....	84
3.3	Materials and Methods	84
3.3.1	Identification of horizontally acquired genes	85
3.3.2	Stratigraphic and donor-recipient analysis	86
3.3.3	Identification and functional analysis of genomic islands.....	88
3.4	Results and Discussion.....	89
3.4.1	Stratigraphic and Donor-Recipient.....	89
3.4.2	Movement of horizontally transferred genes across species	90
3.4.3	Functional analysis of genomic islands.....	92
3.5	Summary	93
Protein and Enzyme Thermostability		95
4.1	Introduction	95
4.1.1	Aims	97
4.2	Materials and Methods	97
4.2.1	Dataset.....	97
4.2.2	Determination of protein thermostability	98
4.2.3	Composition and substitution of amino acid properties	100
4.2.4	Protein structure homology modelling	101
4.2.5	Effect of amino acid substitutions on protein structural stability	103
4.3	Results and Discussion.....	104
4.3.1	Codon usage and the Fariás-Bonato ratio.....	104
4.3.2	Thermostability analysis using minimum folding energy	107
4.3.3	Composition of amino acid properties and residues.....	109
4.3.4	Protein structure homology modelling	113
4.4	Conclusion.....	117
Concluding Discussion.....		120

5.1	Introduction	120
	Bibliography	126
	Appendices	138
6.1	Appendix A	138
6.1.1	Pathway and Breakpoints Results	138
6.2	Appendix B	139
6.2.1	Common Pathway Clustering.....	139
6.3	Appendix C	141
6.3.1	Horizontal gene transfer stratigraphic analysis	141
6.4	Appendix D	150
6.4.1	Thermostability analysis of orthologous sequences	150
6.5	Appendix D	250
6.5.1	Distribution of amino acid residue and property analysis	250

Table of Figures

Figure 1.5.1: Phylogenetic relationship of <i>Deinococcus-Thermus</i>	7
Figure 1.7.1: RNA secondary structure prediction from a sequence.....	29
Figure 1.7.2: Components of an RNA secondary structure.....	29
Figure 1.8.1: G+C content, difference, codon bias for <i>Helicobacter pylori</i> (Karlin <i>et al.</i> , 1998a).....	36
Figure 1.8.2: Codon Adaptation Index against χ^2 for <i>E. coli</i> (Lawrence and Ochman, 1997)	37
Figure 1.8.3: DNA transport machinery in <i>Thermus thermophilus</i> HB27.	39
Figure 1.9.1: The Patchwork Model for pathway evolution Rison and Thornton (2002).....	42
Figure 1.9.2: Software tools pathologic program for constructing PGDBs which includes prediction of pathways and operons	46
Figure 1.9.3: Illustration of prediction of operons in Pathway Tools.....	46
Figure 2.4.1: Identification of <i>M. silvanus</i> DSM 9946 as a reference genome for breakpoint analysis.....	60
Figure 2.4.2: Computation of breakpoints between the reference and the target genomes.	61
Figure 2.4.3: Breakpoints occurring within the same operon; not splitting operons.....	63
Figure 2.4.4: Breakpoints that splits operons. They occur in the same operon in the reference	63
Figure 2.5.1: Genome-metabolic network construction	67
Figure 2.6.1: Distribution of functionally related genes in genomes normalised over expectation	69
Figure 2.7.1: Frequencies of breakpoint events.....	70
Figure 2.7.2: Average operon sizes calculated in the reference and target genomes	71
Figure 2.7.3: Breakpoints occurring inside operons, genomic islands and pathways	72
Figure 2.7.4: Breakpoints occurring within operons or externally	73
Figure 2.7.5: <i>T. scotoductus</i> SA-01 genome rearrangements hotspots with respect to <i>M. silvanus</i> DSM 9946.....	74
Figure 2.7.6: <i>T. thermophilus</i> HB8 breakpoints hotspots with reference to <i>M. silvanus</i> DSM 9946.....	75
Figure 2.7.7: <i>T. thermophilus</i> HB27 breakpoints hotspots with respect to <i>M. silvanus</i> DSM 9946	75
Figure 2.7.8: The logo of sequence overrepresented in vicinities of genome rearrangement breakpoints ..	76
Figure 2.8.1: Clustering of common metabolic pathways using predicted operons.....	77
Figure 2.8.2: Clustering of common metabolic pathways using genomic distance	77
Figure 3.3.1: Stratigraphic donor-recipient illustration (Bezuidt <i>et al.</i> , 2011).....	88
Figure 3.4.1: Age of genomic islands and relative time of insertion.....	91
Figure 3.4.2: Acquisition and movement of genomic islands across species.....	91
Figure 4.2.1: Protein homology modelling process.....	103
Figure 4.3.1: Relationship between codon usage and the (E+K)/(Q+H) ratio in genomes	105
Figure 4.3.2: Codon usage in mesophiles, thermophiles and hyperthermophiles against (E+K)/(Q+H) ratio	106
Figure 4.3.3: Thermostability of genes measured using the (E+K)/(Q+H) ratio	107
Figure 4.3.4: minimum folding energy for 5% groups of genes	107
Figure 4.3.5: Composition of amino acid residues between <i>T. thermophilus</i> HB27 sequences and <i>T. scotoductus</i> SA-01 sequences.	110
Figure 4.3.6: Differences in amino acid property composition.....	111
Figure 4.3.7: Comparison of thermophilic and mesophilic protein structures	117

Table of Tables

Table 1.5.1: Genomic features of completely sequenced <i>T. scotoductus</i> SA-01, <i>T. thermophilus</i> HB27 and <i>T. thermophilus</i> HB8	9
Table 2.1.1: Genome rearrangement events	51
Table 2.3.1: Data set for gene distribution analysis	56
Table 2.6.1: Observed distribution of pairs of functionally related genes for <i>T. scotoductus</i> SA-01 (SA-01), <i>T. thermophilus</i> HB8 and HB27, <i>B. subtilis</i> 168, <i>M. silvanus</i> and <i>ruber</i>	68
Table 2.6.2: Expected distribution of pairs of genes on a chromosome for <i>T. scotoductus</i> SA-01 (SA-01), <i>T. thermophilus</i> HB8 and HB27, <i>B. subtilis</i> 168, <i>M. silvanus</i> and <i>ruber</i>	68
Table 2.7.1: p-values for Wilcoxon rank-sum test for operon length between reference and target genomes	72
Table 2.7.2: Oligomer frequency in coding and non coding sequences of <i>Thermus</i> genomes.	76
Table 2.8.1: Pairwise Wilcoxon <i>t-test</i> , p-values results for common pathways clustering based on predicted operons	78
Table 2.8.2: Pairwise Wilcoxon <i>t-test</i> , p-values results for common pathways clustering based on genomic distance association	78
Table 3.4.1: Analysis of genomic island (GIs) as predicted by SeqWord Snifer for <i>T. scotoductus</i> SA-01, <i>T. thermophilus</i> HB8 and HB27	92
Table 4.2.1: Genomes analysed for codon usage and the (E+K)/(Q+H) ratio classified according to their optimum growth temperature	98
Table 4.3.1: Pairwise substitutions of amino acids in orthologous proteins of thermotolerant <i>T. scotoductus</i> SA-01 and thermophilic <i>T. thermophilus</i> HB27.	112
Table 4.3.2: Matched orthologs for homology modelling	114
Table 4.3.3: Impacts of predominant amino acid substitutions on protein stability predicted by FoldX...116	116
Table 6.1.1: Number of breakpoints affecting pathways	138
Table 6.2.1: Clustering of thirty-eight common pathways	139
Table 6.3.1: Tree topologies for orthologous genes	141
Table 6.4.1: MFE values and resedu substitution statistics calculated for orthologous genes of <i>T. scotoductus</i> SA-01 and <i>T. thermophilus</i> HB27.	150
Table 6.4.2: MFE values and resedu substitution statistics calculated for orthologous genes of <i>T. scotoductus</i> SA-01 and <i>T. thermophilus</i> HB8.	201
Table 6.5.1: <i>p</i> -values for <i>t-test</i> and Wilcoxon <i>t-test</i> (*) analysis of distribution of amino acid residues and properties of orthologous sequences of <i>T. scotoductus</i> SA-01 and <i>T. thermophilus</i> HB27.	250

List of Acronyms

AIC	Artificial Intelligence Center
ANOVA	Analysis of Variance
ATP	Adenosine Triphosphate
BioCyc	Encyclopedia of Biochemical Pathways
BLAST	Basic Local Alignment Search Tool
BLASTp	Protein Basic Local Alignment Search Tool
BsubCyc	Encyclopedia of <i>Bacillus subtilis</i> metabolic pathways
CAI	Codon Adaptation Index
CI	Clustering Index
dNTP	Deoxyribonucleotide Triphosphate
DNA	Deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EcoCyc	Encyclopedia of <i>E. coli</i> metabolic pathways
GI	Genomic Islands
HB27	<i>Thermus thermophilus</i> HB27
HB8	<i>Thermus thermophilus</i> HB8
HGT	Horizontal Gene Transfer
KEGG	Kyoto Encyclopedia of Gene and Genomes
LGT	Lateral Gene Transfer
MetaCyc	Encyclopedia of Metabolic Pathways
MFE	Minimum Folding Energy
NCBI	National Center for Biotechnology Information
OGT	Optimum Growth Temperature
ORF	Open Reading Frame
OU	Oligonucleotide Pattern
PCR	Polymerase Chain Reaction
PGDB	Pathway Genome Database
PS	Pattern Skew
rRNA	Ribosomal Ribonucleic Acid
RNAFold	Ribonucleic Acid Folding Algorithm
RV	Relative Variance
SA-01	<i>Thermus scotoeductus</i> SA-01
SRI	Stanford Research Institute
tRNA	Transfer Ribonucleic Acid
TIGR	The Institute for Genomic Research
UNAFold	Unified Nucleic Acid Folding Algorithm

Chapter One

Introduction

1.1 Introduction

Bacteria of the genus *Thermus* inhabit natural and man-made thermal environments such as hot springs, deep mines, compost manure, sewage sludge and domestic hot water (Oshima & Imahori, 1974; Munster *et al.*, 1986; Brock & Freeze, 1969). *Thermus* bacteria are of major interest in this work because of their enzymes which function at higher temperatures. Some *Thermus* species are capable of reducing heavy metals and switching between aerobic and anaerobic respiration under certain conditions which are very essential for industrial application. Enzymes from thermophilic organisms exhibit higher activity and stability than most synthetic and mesophilic enzymes currently been used in industry for production of food, laundry detergent, drugs, DNA replication, clothing, shoe and paper (Lioliou *et al.*, 2004; Lasa & Berenguer, 1993). Thermophilic enzymes resist denaturing due to extreme acidity or alkaline conditions (Rocha *et al.*, 2000; Niehaus *et al.*, 1999). Examples of industrial important thermostable enzymes include proteases, enzymes for the conversion of carbohydrates such as starch, cellulose, xylan, lipases, ligases, DNA polymerase for DNA amplification, ribonucleases, bioleaching enzymes and many more which make strains from the genus *Thermus* excellent candidates for biotechnological application (Lioliou *et al.*, 2004).

Heavy metals such as *Fe(III)*, *Cr(VI)*, *Mn(IV)*, *U(VI)* and *Co(III)* have been found to be reduced by some *Thermus* strains (Balkwill *et al.*, 2004; Opperman & Heerden, 2007). Reduction of *Fe(III)* and *Mn(IV)* among other heavy metals can be applied in biotechnology industry to eradicate heavy metal pollution; control global warming; remove organic contaminants in ground water; flux phosphate and to clear clogged wells among many other uses (Lovley, 1991). High levels of *Cr(IV)* in the human body is carcinogenic, hence its reduction eliminates toxicity in food and air for human health (Rowbotham *et al.*, 2010). *Fe(III)* reduction under anaerobic conditions diverts electrons away from methane producers thereby reducing global methane fluxes into the atmosphere which lowers global warming. Since thermophilic bacteria are able to switch between aerobic and anaerobic respiration, they yield higher productivity of bio-fuels such as bio-diesel, bio-ethanol and bio-methane under anaerobic conditions when temperature rises due to bio-degradation in bio-fuel cells.

Extreme levels of the biosynthetic and enzymatic activities of thermophilic bacteria have been explained by their extraordinary traits of high frequencies of natural transformations as a

dominant mode for horizontal gene transfer. Through this mechanism, bacteria take up free DNA from environments to incorporate it into their genome. Natural transformations permit the transportation of DNA through bacterial membranes and represent a dominant mode for transferring of genetic material between bacteria of distant evolutionary lineages even between members of different domains (Schwarzenlander *et al.*, 2009). Horizontal gene transfer is one of the dominant players through which micro-organisms such as *Thermus* bacteria are able to acquire foreign genes into their genomes through mechanisms such as conjugation, transduction and natural transformation. The newly acquired genes may contribute new functionalities to the micro-organisms for adaptation and survival in hostile environments such as extreme temperature, high or low acidity and higher salt concentrates among many others.

Horizontal gene transfer is expected to be one of the major factors that affect the distribution and organisation of genes and operons on the chromosome. As genes get acquired or lost, they are shifted around as operons get split or merged which in turn affects the coherence of the metabolic network. Ancestral metabolic networks are believed to be highly connected but they disintegrate with development of new and specialised metabolic abilities which are facilitated by the elongation, splitting and duplication of genes and operons as a result of rearrangements. Functionally related genes in bacteria are grouped into operons or co-localised on the chromosome and usually co-expressed (Yin *et al.*, 2010). Studies have reported that co-localised and co-expressed genes mostly belong to the same metabolic pathways and that biological networks are built on modules that perform particular functions and are reused in an evolutionary manner (Spirin *et al.*, 2006). The study of metabolic networks is applied in functional annotation to identify missing enzymes, predict operons and map them onto the metabolic pathways to identify parallel modules which are basically proteins acting on the same biochemical reactions using different substrates (Spirin *et al.*, 2006). The coherence of metabolic networks is related to speciation and the evolutionary pressures which micro-organisms experience.

The rest of this chapter reviews horizontal gene transfer, genome rearrangements, metabolic networks, gene distribution, and protein thermostability in greater detail. *Thermus* bacteria and industrial application of thermostable enzymes are respectively discussed in sections 1.5 and 1.6. Protein and enzyme thermostability and approaches currently being used to predict thermostability are explored in section 1.7. Section 1.8 covers genome rearrangements and horizontal gene transfer. Evolution, prediction and construction of metabolic networks are discussed in section 1.9. A summary of this chapter together with the organisation of the thesis is outlined in section 1.10. Section 1.2 follows with a detailed background to the research problem.

Sections 1.3 and 1.4 briefly present the aims, objectives and the methodology of the study.

1.2 Research Problem

The ability of thermophilic micro-organisms to live in high temperature, extreme *pH*, high salt concentrates and pressure has fascinated both microbiologists and biochemists for many years. The hyperplasticity of *Thermus* bacteria due to its extraordinary natural transformation which facilitates acquiring of genes that encode for thermostable enzymes; the formation of its metabolic pathways and the regulation, the transcription and translation of their thermostable proteins remain unclear. Introduction of new genes due to horizontal gene transfer through natural transformations is theorised to result into genome rearrangements which affect the order of genes on the chromosome. Due to rearrangements, genes could be shuffled on the chromosome and operons could be split or merged; open reading frames could be altered which could result in loss of genes. All these consequently affect the coherence of the metabolic network due to the close relationship that exists between the order of genes on the chromosome and the level of connectedness of the metabolic network. Although higher levels of natural transformations have been reported in thermophilic bacteria as a survival technique (Schwarzenlander *et al.*, 2009), there has been no comparative genomics studies that investigated the level of genome rearrangements and their impact on the metabolic network.

Thermostability is a complex biological phenomenon achieved by versatile mechanisms which also require to be understood specifically the factors that enhance it in *Thermus* species. Studies have come up with conflicting results on factors that enhance protein thermostability in thermophilic organisms. Hence, there is need to specifically investigate these factors in *Thermus* species. Experimental determination of protein thermostability and possible biotechnological application are both time consuming and costly. There is need for *in silico* approaches to efficiently elucidate factors enhancing protein thermostability in *Thermus* species for potential biotechnological application.

In silico studies require huge amounts of data to perform comparative genomics studies and elucidate properties that enhance thermostability. Although several other thermophilic bacteria from other genera have been sequenced, only two strains of *Thermus thermophilus*, HB8 and HB27 have been completely sequenced within the genus *Thermus*. This has been a major limitation for comprehensive comparative genomic studies to elucidate enhancing factors in *Thermus* species. The isolation of *Thermus scotoductus* SA-01 and completion of its genome sequencing and annotation in February 2010 by respectively the Department of Microbial, Biochemical and Food Biotechnology at the University of the Free State in South Africa and the

J. Craig Venter Institute motivated this study. It provided an opportunity for further comprehensive comparative genomic analysis with the aim of gaining deeper understanding at both genomic and metabolic levels, in addition to thermostability of *Thermus* species. It was until the completion of the whole genome sequencing of these strains that it was possible to perform a comprehensive comparative study of *Thermus* species. It was envisaged at the time that an *in silico* approach that would exploit thermophilic properties among *Thermus* species would be ideal to efficiently identify thermostable enzymes for biotechnology application. This study investigates the following research questions:

- Does horizontal gene transfer affect rearrangements of genes on the chromosome?
- Is there a difference in the distribution of functionally related genes on the chromosomes between mesophilic and thermophilic species due to rearrangements?
- To what extent does the rearrangement of genes on the chromosome affect the coherence of metabolic networks?
- Is there any relationship between the distribution of genes, the organisation of metabolic network and levels of thermostability in *Thermus* species?
- What are the factors that enhance protein thermostability in *Thermus* bacteria?

1.3 Aims and Objectives

This study conducted a comparative genome analysis of horizontal gene transfer, genome rearrangements, distribution of genes on the chromosomes and the evolution of metabolic networks. Horizontal gene transfer and genome rearrangements were analysed among *Thermus* species only. Distribution of genes and evolution of metabolic networks were investigated between mesophilic and thermophilic bacteria. This work was motivated by the extraordinary phenomenon of *Thermus* genome plasticity and their role in acquiring genes encoding thermostable enzymes. The acquired genes contribute to the evolution of metabolic pathways and emergence of new and specialised metabolic abilities for survival in thermophilic environments. Currently, gene regulation, protein transcription and translation in thermophilic bacteria are not well understood. In addition, there has been no *in silico* approach developed to determine factors that enhance thermostability for the efficient prediction of thermostable enzymes. There are two main aims of this study:

1. The first aim was to determine the contribution of evolutionary processes such as horizontal gene transfer, genome rearrangements and selective mutagenesis to the speciation of the of *Thermus* bacteria. Under this aim the following objectives were investigated.
 - Determine levels of genome rearrangement due to natural transformation which

included determining the number of disrupted operons and pathways.

- Examine the extent to which genome rearrangements affect the distribution of functionally related genes on the chromosome.
- Investigate the effect of genome rearrangements on the coherence and structure of the metabolic network.
- Identify genomic islands, possible donors, recipient and age of genomic islands, the contribution of horizontal gene transfer to rearrangements and evolution of species in general.

2. The second aim was to identify factors that enhance protein thermostability that could enable the identification of thermostable enzymes for biotechnology application in *Thermus* species. Comparative analysis of levels of thermostability was performed in order to identify factors that enhance thermostability. These were achieved through the following objectives:

- Determine dominant amino acid substitutions for residues and properties that enhance protein thermostability of *Thermus* bacteria.
- Investigate implications of mutational differences that enhance thermostability at protein structure level.

1.4 Methodology

High level of plasticity due to natural transformations in *Thermus* species are considered to be a major survival technique in thermal environments. Natural transformation is one of the main mechanisms for horizontal gene transfer - the acquisition of genetic material from other organisms which form genomic islands on chromosomes. Genomic islands are regions on a chromosome with foreign genetic material whose DNA composition significantly differs from that of the native chromosome. Genomic islands were identified using the combined usage of several programs such as: SeqWord Genome Browser (Ganesan *et al.*, 2008); SeqWord Snifer (Bezuidt *et al.*, 2009); a web portal Island Viewer that combines the prediction results of three algorithms: IslanPick, SIGIHMM and IslandPath DIMOB SIGIHMM (Waack *et al.*, 2006) and IslandPath, DIMOB (Hsiao *et al.*, 2003). The acquisition of new genetic material results in genome rearrangements as chunks of DNA are inserted, deleted, reversed or relocated on a chromosome.

Genome rearrangements were determined using breakpoint analysis among three *Thermus* genome namely: *Thermus scotoductus* SA-01, *Thermus thermophilus* HB27 and *Thermus thermophilus* HB8. Breakpoints were identified by counting relocation events which were the disruption of synteny of homologous genes or genomes between the target and reference

genomes. Homologous genes between genomes that were compared in the synteny were identified using the BLAST program. This was determined by comparing the synteny of genes against a phylogenetically determined ancestral genome of *Meiothermus silvanus* DSM 9946. Breakpoints provided an indication of both the rate and number of rearrangements that occurred between related genomes. Genome rearrangements affected the distribution of genes as they shifted around in the chromosome.

The distribution of genes was affected as a result of the effect of rearrangements occurring on the chromosomal. The difference in distribution of genes was assessed by comparing the average number of functionally related genes against the random distribution of genes in various genomic distance categories. The genomic distance was measured in number of nucleotides. The categories were as follows: 0 to 100; 100 to 1000; 1000 to 10, 000; 10,000 to 100, 000 and 100, 000 to 1,000, 000 bases. Genomic distance computation between two genes took into consideration the orientation of genes and the circularity of the chromosome. The distribution of functionally related genes has a direct effect on the structure and coherence of the metabolic network.

The structure of the metabolic network provides an indication of evolutionary pressures experienced by micro-organisms and the extent to which organisms evolved. In order to measure the coherence of the metabolic network as affected by genome rearrangements, a cross-clustering index also known as a co-efficient was used. It was computed to measure the clustering of metabolic networks. Pathway Tools software was used to reconstruct metabolic networks to provide information about genes involved in catalysing various pathways. A metabolic network is modelled as a graph with enzymes as nodes with two types of edges. A metabolic edge is when enzymes are encoded by genes from the same operon on the chromosome and a genomic edge joins enzymes belonging to the same pathway. The cross-clustering index is basically the ratio of the total number of edges between enzymes belonging to the same operon (genomic edges) in the same pathway to the total number of edges for all enzymes in the pathway.

To measure protein thermostability, an extended UNAFold algorithm was implemented in Python programming language and was run on all coding sequences in a genome. The basic implementation was command based which took one sequence at a time as input. The extended implementation allowed for computation of all sequences in the genome in fasta or genbank formats at once. Orthologous protein sequences from *Thermus thermophilus* HB27 and *Thermus scotoductus* SA-01 with the largest minimum folding energy difference were compared for dominant mutations that affect thermostability. Analysis was done on closely related species to eliminate differences that exist between bacteria of distant taxa but not necessarily due to

differences in thermostability. Mutations that correlated to changes in energy levels were elucidated. These were further analysed at protein structure level to identify their locations and possible structural differences.

1.5 *Thermus* Bacteria

Thermus bacteria belong to *Deinococcus-Thermus* phylum, *Deinococci* class is in the family of *Thermales* and *Thermaceae* according to the NCBI taxonomic classification. The study of *Thermus* bacteria dates back to late 1960s when Brock and Freeze (1969) isolated *Thermus aquaticus* from the hot springs of Yellow Stone National Park in USA. Prior to that, in the hot springs of Japan, Oshima and Imahori (1974) had already isolated *Flavobacterium thermophilum* in 1968 which was later characterized and reported as *Thermus thermophilus* in 1971. Since then, several strains have been isolated in different parts of the world such as Belgium, Britain, Czech Republic, Iceland, Mexico, New Zealand, Portugal, Russia and South Africa (Munster *et al.*, 1986; Williams *et al.*, 1995).

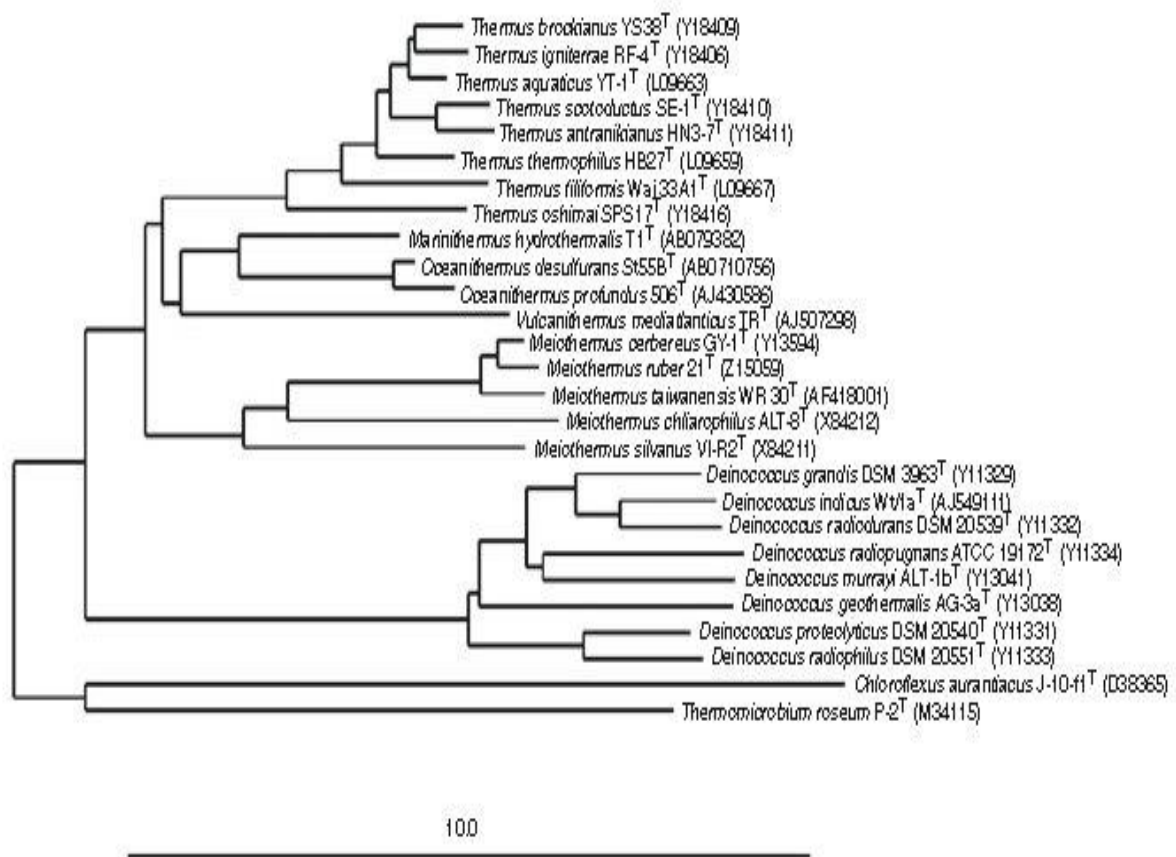


Figure 1.5.1: Phylogenetic relationship of *Deinococcus-Thermus* (Costa *et al.*, 2006)

There are over fifty *Thermus* species but only eight are well known and these are: *T. aquaticus*, *T. thermophilus*, *T. antranikianii*, *T. brokianus*, *T. fliformis*, *T. igniterrae*, *T. oshimai* and *T. scotoductus* (Kurosowa *et al.*, 2005). Out of these, whole genome sequencing has been completed in only two strains of *Thermus thermophilus* (*T. thermophilus*), HB8 and HB27 and recently in *Thermus scotoductus* (*T. scotoductus*) strain SA-01 (Henne *et al.*, 2004; Kieft *et al.*, 1999). The phylogenetic relationship of these *Deinococcus-Thermus* using 16S rRNA sequences as studied by Costa *et al.* (2006) is shown in Figure 1.5.1.

Thermus bacteria inhabits both natural and man-made environments such as hot springs, compost manure, deep mines, sewage and domestic hot water (Brock & Freeze, 1969). The bacteria are rod-shaped, yellow pigmented; gram-negative, non-motile, non-sporulating and heterotrophic (uses organic carbon for growth). They grow both aerobically and anaerobically using nitrate as a terminal electron acceptor. The sizes of the rods range from 0.5 to 0.7 μM in width and 2 to 5 μM in length (Saiki *et al.*, 1972). The yellow pigmentation is due to carotenoid. It is mostly present in species that live in sunlight exposed environments while species found in sunlight deprived environments such as deep mines are colourless as the pigmentation protects cells from sunlight (Pask-Hughes & Williams, 1975). Growth conditions include temperatures between 65 $^{\circ}\text{C}$ and 72 $^{\circ}\text{C}$, which stretches to a minimum of 47 $^{\circ}\text{C}$ and a maximum of 85 $^{\circ}\text{C}$ (Oshima & Imahori, 1974). Their optimal growth *pH* ranges from 5.8 to 8.9 (Kurosowa *et al.*, 2005).

T. thermophilus has three known strains HB8, HB27 and AT62 (Costa *et al.*, 2006) of which the genomes of *T. thermophilus* HB8 and HB27 have already been completely sequenced. This work focused on the two strains of *T. thermophilus*, HB8 (NC_006461) and HB27 (NC_005835). These two strains are comparatively studied with *T. scotoductus* strain SA-01 (NC_014974). The genomes of these three strains have been completely sequenced and are available in the NCBI public database. For the rest of the discussion *T. thermophilus* HB8, *T. thermophilus* HB27 and *T. scotoductus* SA-01 will be interchangeably referred to as HB27, HB8 and SA-01 respectively. The strain HB8 has a genome size of 1,849,742 base pairs (bp) long with two plasmids pTT27 and pTT8. The genome size of HB27 is 1,894,877 bp and has one plasmid pTT27. Further details on their genomic features are provided in Table 1.5.1. These three strains have been studied in comparison with moderately thermophilic bacteria, *Meiothermus ruber* (NC_013946) and *M. silvanus* (CP002042) and mesophilic bacteria *E. coli* (NC_000913) and *Bacillus subtilis* (NC_000964).

Table 1.5.1: Genomic features of completely sequenced *T. scotoductus* SA-01, *T. thermophilus* HB27 and *T. thermophilus* HB8

Features	HB27	HB8	SA-01
Genome Size (bp)	1,894,877	1,849,742	2,346,803
Plasmid(s) Size (bp)	232,605	256,992; 9,322	8,383
Number of Genes	1,980 (224)	1,994 (253; 11)	2,506 (12)
G+C Content (%)	69.43 (69.15)	69.4 (69.3; 69.0)	64.9 (65.0)
Growth Temperature	60 - 85 °C	60 - 85 °C	60 - 65 °C

As shown in Table 1.5.1, *T. scotoductus* SA-01 strain has a genome of 2,346,803 base pairs with one plasmid known as pTT8. The bacteria were isolated from a gold mine in South Africa at Witwatersrand at a depth of 3.2 km with 60 °C temperature. The gold mine is a 2.9 billion year old formation that share minor volcanic units and conglomerates (Kieft *et al.*, 1999). The strain SA-01 is non-pigmented as it is found in under the ground in sunlight deprived environments. It grows both aerobically and anaerobically using nitrate as a terminal electron acceptor. It is also known to reduce heavy metals, which has numerous biotechnology benefits. Its whole genome was completely sequenced in February 2010 which provided an opportunity for a comprehensive comparative genomics study. Other known strains for *T. scotoductus* were isolated in Mexico, United Kingdom, Iceland and Portugal are X1, NMX2 A.1, VI7, NH, DI and SE1 (Balkwill *et al.*, 2004; Skirnisdottir *et al.*, 2001; Costa *et al.*, 2006). *T. scotoductus* X1 was isolated from hydrothermal hot tap water in Iceland. The strains NH and DI for *T. scotoductus* are nonpigmented and were isolated from hot tap water in London.

1.6 Industrial Application

Bacteria have many properties that are further enriched by their diversity which can be useful for industrial application. The application of micro-organisms in industry has huge economic benefits. In various industrial sectors they have enhanced production and eventual profits due to reduced costs. They have facilitated production of additional products from industrial wastes which does not only reduce pollution, but also allows maximum utilisation of inputs and raw materials. Industrial application of micro-organisms involves manufacturing of classical industrial chemicals or using microbial living cells to synthesise biological chemicals. Industrial chemicals are produced by microbes as by-products when inoculated on a specific medium. For bio-remedial purposes for example, bacteria can be inoculated on polluted environments such as land, air or water where they eradicate pollution through their metabolic processes. Instead of directly inoculating the organisms, their enzymes can be extracted or artificially synthesised by reengineering their thermostability enhancing properties.

The industrial application of micro-organisms is discussed in two parts. The use of micro-organisms for heavy metal reduction is discussed first. Then, industrial application and use of various thermostable enzymes is explored. Heavy metal reduction is useful for bio-mining, bio-remediation, controlling global warming and reducing toxicity in food for human health. Microbial enzymes are useful among many industrial applications for production of food, paper, plastics, shoes, laundry detergent, bio-fuels and pharmaceuticals.

1.6.1 Heavy metal reduction

Several metals are reduced by micro-organisms from one form to the other, usually from a harmful to a harmless or less beneficial to more beneficial form. For example, *Fe(III)* can be reduced to *Fe(II)*; *Cr(VI)* which is extremely carcinogenic to *Cr(V)*, and *Mn(IV)* to *Mn(II)*. *Thermus scotoductus* SA-01 reduces heavy metals such iron *Fe(III)*, zinc (Zn), manganese *Mn(IV)*, chromium *Cr(IV)*, copper *Co(III)* EDTA, uranium *U(VI)* and sulphur (*S*) by using them as terminal electron acceptors for growth under both aerobic and anaerobic conditions. Kieft *et al.* (1999) found that *Thermus scotoductus* strains SA-01 and NMX2 A.1 grew in basal medium amended with lactate and any other electron acceptor such as oxygen, nitrate and *Fe(III)*-NTA but not in the absence of an electron acceptor. The strains were able to reduce *Fe(III)*-NTA to *Fe(II)* with the disappearance of lactate in correlation with cell growth. Iron reduction and cell growth proceeded even without lactate, however levels of *Fe(II)* produced and cell growth yields were significantly lower. Similar reduction results were observed in *Mn(IV)*, *Co(III)* EDTA, *Cr(IV)* and *U(V)*. Opperman and Heerden (2007) found that *T. scotoductus* SA-01 was able to reduce *Cr(VI)* aerobically when grown in a complex organic medium containing *Cr(VI)* concentration with up to 0.5 mmol l^{-1} . *Cr(VI)* reduction results were observed under non growth conditions both with and without electron donors at optimum temperature of 80°C with pH of 7. It was found that under anaerobic conditions, the strain used a membrane bound mechanism. A secondary enzyme localised in the cytoplasm was used in aerobic environment. *Thermus scotoductus* strain IT7254 isolated in sulphide rich hot springs of Iceland grew and used thiosulfate and elementary sulphur as electron donors with oxygen and nitrate as electron acceptors (Skirnisdottir *et al.*, 2001). Results after two to four days of incubation on $16 \mu\text{M}$ thiosulfate gelrite medium plates at 65°C showed growth with sulphate as a product. The strain was able to oxidise sulphur to sulphate in the presence of acetate but less efficiently with thiosulfate. In liquid thiosulfate medium with no carbon sources, the bacteria grew extremely poorly but precipitated sulphur in 10 days.

The use of micro-organisms has become an important alternative to conventional physical and

chemical treatments for the removal of toxic heavy metals from polluted environments through bio-remediation. Other application of heavy metal reduction include: bio-mining (Anold *et al.*, 1988), controlling global warming (Cicerone & Oremland, 1988), purifying contaminated water (Bostrom *et al.*, 1988), and removing of clogs in pipes and reducing toxicity in food (Rowbotham *et al.*, 2010). There are quite a number of metals that are reduced by microbes and are increasing with ongoing research. However, discussion of the importance of heavy metal reduction in this study is limited to iron *Fe(III)*, manganese *Mn(IV)* and chromium *Cr(VI)* as examples.

1.6.1.1 Iron *Fe(III)* and manganese *Mn(IV)* reduction

1.6.1.1.1 Global warming reduction

It is estimated that 40% of global methane is accounted for shallow fresh water environments such as rice paddies and swamps where anaerobic conditions occur during flooding times (Cicerone & Oremland, 1988). *Fe(III)* reduction is useful for organic matter decomposition as it is oxidised to carbon dioxide as an electron acceptor. This phenomenon has been observed in most aquatic sediments and waterlogged soils under anaerobic conditions (Froelich *et al.*, 1979). *Fe(III)* reduction for organic matter decomposition seem to be more abundant than *Mn(IV)* and sulphur (*S*) reduction; and methane production as *Fe(III)* out-competes them as an electron donor. The advantages include reduction in global methane fluxes into the atmosphere which lowers global warming. *Fe(III)* oxides generated during oxidation period divert electrons away from methane producers depending on the temperature, organic matter and iron content of the soils (Lovley, 1991).

1.6.1.1.2 Removal of organic contaminants in groundwater and release of trace metals

Under anaerobic conditions, contaminated ground water *Fe(III)* becomes the most abundant electron acceptor. *Fe(III)* and *Mn(IV)* are highly insoluble and they clog wells, discolour clothing and impair taste. This problem is solved by applying micro-organisms that reduce *Fe(III)* and *Mn(IV)* to soluble *Fe(II)* and *Mn(II)* respectively (Baedecker & Back, 1979; Lovley *et al.*, 1989).

Fe(III) and *Mn(IV)* microbial reduction play a huge role in fluxing of phosphate and other trace metals into water supplies. *Fe(III)* and *Mn(IV)* oxides in sediments to a larger extent strongly adsorb a wide variety of toxic trace metals which are a potential hazard to the aquatic ecosystems and drinking water supplies. Their reduction to soluble *Fe(II)* and *Mn(II)* purifies the water by releasing these contaminants into the surrounding water (Bostrom *et al.*, 1988).

1.6.1.1.3 Bio-mining and generation of minerals

Microbial reduction of *Fe(III)* has been proposed as a mechanism for extraction of *Fe(III)* from

ore (Anold *et al.*, 1988). *Fe(III)* reducing micro-organisms dissolve it into soluble *Fe(II)* which requires less reduction to elemental iron. Reduction of manganese in ferromanganese concretion and crusts results in the release of dissolved manganese, cobalt, copper and nickel as a microbial leaching technique; although this approach is relatively slow to existing chemical processes (Breithaupt, 2001). Microbial *Fe(III)* reduction can result in the generation of several iron and manganese minerals; important is *Fe(II)* containing minerals in sedimentary environments such as magnetite ($Fe(II)Fe(III)_2O_4$), siderite $FeCO_3$ and vivianite $Fe_3(PO_4)$; and rodochrosite $MnCO_3$, a manganese containing metal (Mumford, 1913; Bell *et al.*, 1987).

1.6.1.2 Chromium *Cr(VI)* reduction

Chromium *Cr(VI)* is an elementary metal that exists in several oxidation states and certain hexavalent compounds are known to be mutagenic and carcinogenic which presents occupational health hazards. It is a naturally occurring element found in significant quantities in the earth's crust, in surface rocks, soil and sea water. It exists in several forms but only trivalent *Cr(III)* and hexavalent, *Cr(VI)* are common in natural environments in the form of chromate (CrO_4^{2-}) and dichromate ($Cr_2O_7^{2-}$) (Opperman & van Heerden, 2008).

Widespread use of chromium in various industrial applications and infrequent incorrect disposal of by-products and wastes have created serious environmental hazards. Chromium as *Cr(III)* and *Cr(VI)* are emitted into the atmosphere from various sources such as public powder, fuel combustion and waste incineration. Other sources include iron and steel production, road transport through wearing of tires and brakes, and releases from photocopiers. It enters aquatic environments from industrial sources through discharged liquid affluent and dumping waste. It contaminates soil largely through wastes disposal activities. Chromium as sewage, iron and steel industrial and household wastes are deposited to landfills and farmlands. The general population get exposed to chromium through food, drinking water, soil in young children and air as a result of inhalation, ingestion and dermal contact (Canada, 1994; Rowbotham *et al.*, 2010).

Cr(III) is an essential nutrient that is usually ingested in insoluble form and has very low degree of absorption from the gastrointestinal tract. Even when absorption of ingested *Cr(VI)* is slightly higher, the health risks posed are low because acidic conditions in the stomach reduce it to *Cr(III)*. It can also be reduced in terminal airways in the epithelial lung fluid and pulmonary alveolar macrophages. However, there is evidence that chromium in form of *Cr(III)* may accumulate in human tissue. Ingested *Cr(VI)* can be reduced by human acidity in the gut to *Cr(III)* which can easily be absorbed and distributed to parts of the body to transverse cell

membranes by diffusion. When $Cr(VI)$ has not been efficiently reduced by the gut or after inhalation it can be directly absorbed and taken by cells. In the cell it is reduced to $Cr(III)$ through formation of reactive intermediate $Cr(V)$ and $Cr(IV)$ which reacts with intracellular components producing inorganic complexes, thus essentially trapping the metal ions for the duration of the life of the cell. Acute toxicity of $Cr(III)$ and $Cr(IV)$ in experimental animals and humans through oral, inhalation and dermal contact had several health effects. In humans, oral uptake of chromium results in gastrointestinal bleeding, fluid loss and death; inhalation results in nasal ulceration, septum perforation, nasal irritation, haemoptysis, asthma and bronchitis, dermal toxicity results in nausea, shock, coma and death. Effects in rats and mice exposed to higher dosage include gastric ulceration, pulmonary congestion, gastrointestinal edema, erosion and discoloration of gastric mucosa, cyanosis, tail necrosis and decrease in spleen and liver size (Rowbotham *et al.*, 2010).

1.6.2 Thermostable enzymes

Both mesophilic and thermophilic micro-organisms have enzymes useful for industrial application. The use of enzymes from mesophilic organisms has an additional cooling cost in order to maintain them at their low activity temperature. Industrial processes at low temperatures bring in additional challenges. Solubility is usually low due to high viscosity. Since most micro-organisms have low optimum growth temperature, there is high probability of contamination on the end product. Hence, in addition to cost, thermal sensitivity is one of the major challenges for industrial application of mesophilic enzymes.

Thermophilic enzymes are inherently more stable and active to heat, resistant to organic solvents, detergents and to extreme pH than their mesophilic homologs which makes them more useful for industrial application (Leuschner & Antranikian, 1995; Lioliou *et al.*, 2004; Lasa & Berenguer, 1993). They facilitate industrial processes at high temperatures without denaturing. This reduces the risk of contamination by common mesophiles which are abundant in low temperatures. Most compounds are bio-available and soluble at high temperatures where there is higher reaction rate due to the decrease in viscosity and increase in solubility (Becker *et al.*, 1997; Lasa & Berenguer, 1993; Kristjansson, 1989). The need for heavy cooling system is eliminated thereby reducing cooling costs. Since, reaction rates accelerate at high temperatures, delays due to cooling and circulation to increase solubility are eliminated hence, production rates and eventual profits increase.

Several enzymes have been identified from thermophiles for biotechnology application. These include: amylases, pullulanases, cellulases, chitinases, xylanases, pectinases, isomaerases,

esterases, and DNA replication enzymes. These enzymes are applied in various industrial sectors such: laundry, starch, dairy, textiles, brewing, baking, leather; pulp and paper; chemical and pharmaceutical industries, and environmental biotechnology. A discussion covering all the enzymes is not feasible in this work. However, the next section proceeds to discuss some of these enzymes and their industrial application.

1.6.2.1 Starch degradation

Degradation of starch is applied in food and bio-refining industry. It is used in manufacturing of bread where starch is broken down into flour which is converted to glucose, fructose and other syrups. In brewing, polysaccharides are broken down to malt for the manufacturing of beer. Amylases are used in textiles to remove starch from woven fabric which is applied to threads as adhesives to prevent damage. It is applied in the laundry industry to remove resistant starch residues. Glucose from starch can be fermented further to produce bio-fuels such as bio-diesel, bio-ethanol and bio-methanol (Haki & Rakshit, 2003).

The complex structure of starch requires a number of enzymes to be applied in degradation steps. Starch degradation enzymes are categorised into three groups which are α -amylases, β -amylases and pullulanases. The degradation process itself occurs in two steps. In the first step, starch is solubilised at 110 °C with α -amylases commonly extracted from *Bacillus subtilis* and *Bacillus amyololiquefaciens*. In the second step, meltodextrins are digested into glucose and oligosaccharides by type II enzymes (Lasa & Berenguer, 1993). Starch comprise α -glucose linked by α -1,4- or α -1,6-glycosidic bonds to form amylose, a linear polymer consisting of α -1,4-linked glucopyranose residues and amylopectin a branched polymer containing a α -1,6-glycosidic linkages at branching points. A number of enzymes are needed for its degradation which are grouped into endoacting (α -amylase) and exoacting (glucoamylases, α -glucosidases and β -amylases) enzymes (Niehaus *et al.*, 1999). Endoacting enzymes hydrolyse linkages in the interior of the starch polymer while the exoacting enzymes attack the substrate from the non-reducing end producing oligo and monosaccharides. Debranching enzymes also known as pullulanases hydrolyse α -1,6-glycosidic bonds which are essentially categorised on the basis of their substrate specificity into two types which are *type I* and *type II*. *Type I* hydrolyses α -1, 6-linkages in pullulan and branched oligosaccharides and *type II* attack both α -1,6-glycosidic linkages in pullulan and α -1,4-linkages in other oligosaccharides and polysaccharides.

1.6.2.2 Cellulases

Cellulose is an abundant raw material produced in large quantities from forest and agriculture

wastes. It consists of glucose units linked by β -1, 4-glycosidic bonds in a linear fashion and is the most abundant source of chemicals, fuels and feed. It is highly resistant to digestion and hydrolysis. It requires enzymes such as endoglucanases, exoglucanases and β -glucanases. Thermostable cellulase, an enzyme that works on cellulose has been extracted from several micro-organisms such as *Anaerocellu thermophilum*, *Bacillus subtilis*, *Pyrococcus furiosus*, *Pyrococcus horicoshi*, *Rhodothermus marinus*, *Thermotoga maritima* MSB8 and *Thermotoga neapolitana*. Cellulase is an endoglucanases that randomly hydrolyse cellulose to produce oligosaccharides, cellobiose and glucose. Exoglucanase hydrolyse β -1, 4-D-glucosidic linkages in cellulose releasing cellobiose from the non-reducing end, while β -glycosidases hydrolyse cellobiose to glucose (Matsui *et al.*, 2000).

Efforts are underway in the food industry to produce sugars from cellulose vegetal wastes. Other industrial applications include: paper bleaching for recycling, colour extraction for juices, as detergents for colour brightening and softening, bio-stoning of jeans, pre-treatment of biomass that contain cellulose to improve nutritional quality and in pre-treatment of industrial waste (Lasa & Berenguer, 1993; Niehaus *et al.*, 1999). Cellulose is also useful in laundry as a textile detergent to remove cotton fuzz which accumulates with excessive washing and for bio-polishing processes of cotton. Cellulolytic biomass can also be applied to produce bio-ethanol.

1.6.2.3 Xylanases

Xylan is another abundant organic substance dominant in main polymeric compound of hemicelluloses which mainly forms the plant cell wall, a main reservoir of fixed carbon in nature. Xylose residues linked by β -1, 4-glycosidic bonds form the main component of the hyper-polymer. Several enzymes are required for its complete degradation. β -1, 4-xyllans and xylobiose are hydrolysed by removing successive xylose residues from the non-reducing termini by β -1,4-xylosidase or β -xylosidase or β -1,4-xylan xylohydrolase or xylobiase or exo- β -1,4-xylosidase (EC 3.2.1.37). β -1, 4-xylosydic linkages in xylans are hydrolysed by the endo- β -1,4-xylanase (EC 3.2.1.8), or β -1, 4-xylan xylanohydrolases. The non-reducing terminal α -L-arabinofuranoside residues in applicable in α -L-arabinosides are hydrolysed by α -Arabinofuranosidase or arabinodase (EC 3.2.1.55). β -1, 4-xylosyl linkages in some glucuronoarabinoxylans are attacked by glucuronoarabixylan endo β -1, 4-xylanase or feraxan endoxylanase or glucuroarabinoxylan β -1,4-xylanohydrolase (EC 3.2.1.136). The acetyl group from xylan are removed by acetylxylanesterase (EC 3.1.1.6) (Niehaus *et al.*, 1999; Annison, 1992).

These enzymes are applied as food additives in poultry to increase efficiency; in wheat for improving dough handling, extraction of coffee and plant oils, and baking quality and in

production of paper and pulp (Beg *et al.*, 2001). It is one of the fastest growing industries that demand the use of thermostable enzymes. Wood for production of pulp is treated at high temperature and basic *pH* to disrupt cell walls and facilitate the removal of lignin. The enzymes lack cellulolytic activity to avoid hydrolysis of cellulose fibres. Xylanases have low molecular mass to enable diffusion in pulp fibres and also are of lower costs. Commercially available enzymes are currently expensive and are active at lower temperatures. However, thermostable xylanases isolated from a number of bacteria and fungi exhibit higher level of activity (Lasa & Berenguer, 1993; Haki & Rakshit, 2003; Cordeiro *et al.*, 2002). Members of *Bacillus* species, *Streptomyces* species, *Thermoascus aurantiacus*, *Fusarium proliferatum*, *Dictyogloclomus* species and *Thermatogales* species are known to produce thermostable xylanases. There is a need for more micro-organisms producing these thermostable enzymes to meet industrial demand. Furthermore, thermophilic enzymes in this industry are preferred because they eliminate waste products which pollute the environment.

1.6.2.4 Chitinases

After cellulose, chitin is the second most abundant naturally occurring bio-polymer. It consists of a linear β -1, 4-homopolymer of N-acetylglucosamine residue. It is usually attached to polysaccharides and proteins covering layers of insects and the cell walls of fungi, crab, shrimp and prawns. It is reported that over 2.5 million tonnes of waste are produced after consumption of shrimps and prawns. With improved and economically feasible chemical, enzymatic and novel biotechnology techniques, these wastes could be converted into value added products such as chitin and its derivative product chitosan produced by deacetylation reaction in 40% sodium hydroxide solution. Their application include waste water clearing, preparation of cosmetics, paper production, cement, heavy metal chelating agents and for medical and veterinary use (Haki & Rakshit, 2003; Majeti & Kumar, 2000; Shahidi *et al.*, 1999). Enzymes for chitin degradation include endo-acting chitin hydrolase chitinase A (EC 3.2.1.14) and exo-acting hydrolase chitinase B and N-acetyl-D-glucosaminidase (EC 3.2.1.52). Well known sources of chitinases are *Bacillus licheniformis* X7u, *Bacillus* sp. BG11, *Bacillus stearothermophilus*, *Thermococcus chitinophagus* and *Streptomyces thermoviolaceus* OPC-520 (Niehaus *et al.*, 1999).

1.6.2.5 Proteases

Proteases are used to convert proteins into amino acids and peptides. They comprise the largest amount of enzymes produced on commercial scale worldwide than any other biotechnologically used enzyme. They constitute over 65% of the total commercial enzyme industry. They are

categorised into two: exopeptidases that cleave off amino acids from protein chains and endopeptidases which cleave peptide bonds within the protein (Rao *et al.*, 1998). When classified according to their catalytic site nature they fall into serine, cysteine or aspartic proteases or metalloproteases. These enzymes are useful in production of food, pharmaceuticals, detergents, leather and clothing industry. Serine proteases are used as additives to household detergents for laundering to resist denaturing by detergents due to alkaline conditions. Proteases with high keratinolytic and elastolytic activities are used for soaking in leather industry (Miguel *et al.*, 2006; Haki & Rakshit, 2003; Synowiecki, 2010).

A number of heat resistant proteases have been identified in hyperthermophilic archaea belonging to genera *Desulfurococcus*, *Sulfolobus*, *Staphylothermus*, *Thermococcus*, *Pyrobaculum* and *Pyrococcus*. *Bacillus stearothermophilus* and *licheniformis* are some of the few thermophilic bacteria from which proteases have been isolated. Most proteases from extremophiles belong to serine type and are stable at high temperatures and even with high concentration of detergents and denaturing agents (Niehaus *et al.*, 1999).

1.6.2.6 Lipases

Lipases of microbial origin are versatile and bring about a range of reactions including hydrolysis, interesterification, esterification, alcoholysis, acidolysis and aminolysis. The esters produced play a relevant role in food industry as flavour and aroma constituents. Long chain methyl and ethyl of carboxylic acid moieties provide valuable oleochemical species that may function as bio-fuel for diesel engines. Long chains of carboxylic acid and alcohol moieties have applications as lubricants and additive in cosmetic formulations (Jaeger & Reetz, 1998). Other applications include removal of the pitch from pulp produced in the paper industry; hydrolysis of milk fat, production of cheese in the dairy industry; removal of non cellulosic impurities from raw cotton before further processing with dye into finished products. Other applications include drug formulation in pharmaceutical industries, removal of subcutaneous fat in the leather industry and production of bio-diesel from vegetable oil (Miguel *et al.*, 2006; Haki & Rakshit, 2003). Most lipases function at 45 °C, hence enzymes required need to operate at maximum temperatures of 50 °C. However, there are fats with higher melting points and inhibit enzymatic activity. Some seed oils require *pH* of 5.0 and temperatures up to 65 °C. Although lipases are widely spread in the earth flora, they are more abundantly found in bacteria, yeast and fungi with *Bacillus* species being the main source. Lipases from thermophilic *Rhizopus oryzae* and *Bacillus thermoleovorans* strains function moderately at extreme temperatures and *pH*.

1.6.2.7 DNA polymerases

Huge advances have been made in genetic engineering as a result of polymerase chain reaction (PCR) which is catalysed by DNA polymerase (EC 2.7.7.7). The steps in the process involves denaturing or melting DNA strands which separates DNA strands at 90-95 °C; renaturation or primer annealing at between 40 to 75 °C and synthesis of primer extension at 72 °C. In early PCR DNA polymerase isolated from *E. coli* was used but it denatured at elevated temperatures. This required adding new enzymes at each cycle which was expensive and time consuming (Mullis *et al.*, 1986; Saiki *et al.*, 1972). The first thermostable DNA polymerase was Taq polymerase isolated from *Thermus aquaticus*. However, it exhibits activity only from the 5' to 3' and not 3' to 5' end which results in low base fidelity rate as the enzyme fail to correct missincorporated nucleotides. PCR process is impeded with several challenges such as rate of phosphodiester bond formation, binding of dNTP by polymerase, rate of pyrophosphate release, contamination after miss incorporation and 3' 5' exonuclease proof reading capacity of enzymes. Although most of these can be rectified by improved fidelity in PCR, thermostable DNA polymerase with all these characteristic would be more preferred to improve results (Niehaus *et al.*, 1999; Haki & Rakshit, 2003). DNA polymerase have been isolated from *Bacillus stearothermophilus*; *Pyrococcus* species of *P. GB-D*, *P. furiosus*, and *P. woeisi*; *Thermus aquaticus*, *T. fliformis*, *T. favus*, *T. thermophilus*; *Thermotoga maritima* and *Thermotoga litoralis*.

1.6.2.8 Ligases

Construction of sequencing primers through high temperature ligation of hexameric primers, detection of trinucleotide repeats through repeat expansion detection, or DNA detection by circularisation of oligonucleotides are some of the application for thermostable ligase. In heritable diseases, a powerful approach that utilises DNA ligases is applied to detect single base mutations in specific nucleotide sequences (Nilsson *et al.*, 1994; Schalling *et al.*, 1993). A major drawback in the technique which involves two oligonucleotides hybridizing from the 3' to 5' with the DNA ligase linking them covalently is the detection of relatively small amount of the product and high amount of background due to unspecific ligation. These problems were overcome by ligase chain reaction where denaturing, annealing and ligation is repeated several times and performing the reaction near the melting point of the primers.

The first ligase was derived from *Thermus thermophilus* HB8 and it displayed a wide range of activity (Jónsson *et al.*, 1994). Later several thermostable DNA ligases have been discovered

from *Thermus scotoductus* and *Rhodothermus marinus*.

1.7 Protein and Enzyme Thermostability

Thermostability is the resistance of a substance to irreversible chemical and physical change of its structure and property due to increase in temperature. Thermostable proteins preserve their unique chemical and spatial structure of the polypeptide chain under extreme temperature (Zhou *et al.*, 2008). Protein thermostability has several advantages. It is influenced by a number of factors which are discussed in subsections that follow. Enzyme thermostability and protein thermostability have been used interchangeably throughout the discussion.

1.7.1 Factors enhancing protein thermostability

There has been a growing interest to elucidate factors that enhance thermostability of thermophilic proteins. The elucidation of thermostability enhancing factors is important not only to understand protein folding and stability but also for synthesising enzymes which efficiently work in higher temperatures. Identifying and understanding factors that enhance protein thermostability is a long standing problem. Several factors have been ascribed to enhance protein thermostability which include codon usage, amino acid composition, stabilizing sequences, ion pairing, hydrogen bonds, hydrophobic interactions, cavities and core packing, sulphide bridges; inter subunit stabilisation and oligomerization, and increased surface hydrophobicity (Fariás & Bonato, 2002; Fariás & Bonato, 2003; Kumar *et al.*, 2000). Other properties that statistically correlate with thermostability of thermophilic proteins include higher residue volume, more charged amino acids (Glu, Arg, and Lys) and reduced uncharged polar residues (Ser, Thr, Asn, and Gln) (Haney *et al.*, 1999). More results with increase in data demonstrate that properties of side chains of amino acids were determinants of thermostability. Amino acids are classified into four groups according to properties of side chain amino acids: (1) non-polar, (2) polar, uncharged (3) polar charged and (4) aromatics. These are briefly discussed next in relation to protein thermostability.

1.7.1.1 Codon usage

A study by van der Linden and Fariás (2006) showed preferred codon usage of AGR among hyperthermophiles and thermophiles when coding for arginine while mesophiles mostly coded with CGN. They analysed complete genomes of 173 micro-organisms from the TIGR database comprising 10 hyperthermophiles, 10 thermophiles and 153 mesophiles. These were classified according to their optimum growth temperature with mesophiles being below 50 °C, thermophiles

between 50 °C and 80 °C whilst hyperthermophiles had optimum growth temperature above 80 °C. A similar analysis was done by Fariás and Bonato (2002) using 28 complete genomes. The preferred coding with AGR implied positive error minimisation during mutations. This allowed mutations of G to A in the second base resulting into lysine codons which are also positively charged, an important property for thermostability. The choice of CGN in mesophiles increased negative error minimisation resulting in histidine and glutamine codons. They proposed that AGR could be used as a signature for hyperthermophiles and thermophiles (Fariás & Bonato, 2002).

1.7.1.2 Amino acid composition

Studies have established a relationship between protein characteristics and amino acid composition (Kumar *et al.*, 2000; Ponnuswamy *et al.*, 1982; Trivedi *et al.*, 2006). These studies investigated amino acid composition between mesophilic and thermophilic micro-organisms in order to determine their influence on thermostability. Differences in amino acid distribution between mesophilic and thermophilic organisms were highly significant despite high sequence homology and that some of the differences were associated with phylogenetic relationships (Kumar *et al.*, 2000). Correlation was established between melting temperature and amino acid composition of stabilizing and destabilizing group of amino acids (Ponnuswamy *et al.*, 1982). It is interesting to note that although there is a general consensus agreement that protein characteristics are influenced by amino acid composition, there are conflicting results on the actual amino acid composition that distinguish thermophilic from mesophilic proteins. The variations in amino acid composition appear to be not only protein specific, but also organisms and even taxa specific (Trivedi *et al.*, 2006). For instance, isoleucine is preferred in *Methanococcus* and *Picrophilus torridus* while in others valine or glycine is preferred. Lysine and tyrosine is preferred over arginine in *Methanococcus jannaschii*. Ile and alanine are preferred over tyrosine in *Bacillus stearothermophilus* (McDonald *et al.*, 1999).

An increase in glutamic acid and lysine in hyperthermophiles and a decrease in glutamine and histidine was observed by Fariás and Bonato (2003). Based on this observation, the (E+K)/(Q+H) ratio also known as the Fariás-Bonato ratio was devised to classify mesophiles, thermophiles and hyperthermophiles. The ratio was determined to be above 4.5 in hyperthermophiles, less than 2.5 in mesophiles and between 3.2 and 4.6 in thermophiles. Chaperons and ligases did not comply with this classification and were therefore, excluded from the computation. These results were obtained from an analysis of 6 hyperthermophiles, 4 thermophiles and 18 mesophiles. In a similar study, a strong positive correlation was observed between (E+K)/(Q+H) and AGR codon usage (van der Linden & Fariás, 2006). In a study by Argos *et al.*, (1979), glycine, serine, lysine, and

aspartic acid in mesophiles was replaced in thermophiles by alanine, threonine, arginine, and glutamine respectively. Analysis of 373 homologous thermophilic and mesophilic proteins revealed an increased fraction of lysine, arginine, and glutamic acid and a decrease of alanine, aspartic acid, asparagine, glutamine, threonine, serine and histidine in thermophilic proteins. Amino acid distribution is further discussed in the proceeding subsections with respect to whether they are non-polar, polar charged, polar non-charged and aromatic between thermophilic and mesophilic proteins.

1.7.1.3 Non-polar amino acids

Non-polar amino acids (glycine, alanine, valine, leucine, isoleucine, proline and methionine) do not have polar atoms (carbon and hydrogen) in their side chains. The contribution of each of these non-polar amino acids to thermostability is discussed. Hydrophobic non-polar amino acids are slightly more abundant in thermophiles. They generally increase rigidity of proteins due to hydrophobicity (Chakravarty & Varadarajan, 2002). Hydrophobic residues are found in the interior while polar residues occur on the surfaces of globular proteins.

Alanine (with a methyl group) occurs with a lower frequency in exposed state and higher frequency in well buried states of thermophilic proteins (Pack & Yoo, 2005). The short alkyl group interacts resulting in compact packing. This further translates to the fact that Ala is the best helix forming residue in thermophiles. Although thermophilic proteins have lower beta-branched residues (Val, Ile, and Thr) which destabilises the helix formation, closer analysis showed that they have higher isoleucine and valine (Kumar *et al.*, 2000).

Some of the non-polar amino acids are also aliphatic which contributes to hydrophobic interaction, a major force in maintaining inner protein conformation stabilisation. Non-polar aliphatic amino acids among hydrophobic residues include alanine, valine, leucine and isoleucine. The relative volume of a protein occupied by the aliphatic side chains defines the aliphatic index which is a positive factor in increased thermostability in globular proteins. Thermophilic proteins have higher hydropathy and aliphatic index because of higher leucine composition.

Conflicting results have been found about the occurrence of glycine, one of the non-polar amino acids, between thermophilic and mesophilic organisms. Glycine is known to make volume or cavity in the inner parts of protein structure. Panasik *et al.*, 2000 found that thermophilic proteins have fewer glycines in a particular region of the structure while Pack and Yoo (2005) found no typical occurrence of glycine in thermophilic proteins. Another non-polar amino acid, proline makes rigid or turn conformations in protein structures due to its pyrrolidine ring which can only adopt a few configurations (Watanabe *et al.*, 1997). A high occurrence of proline has

been observed by several studies in thermophiles (Xu *et al.*, 2003; Pack & Yoo, 2005; Sadeghi *et al.*, 2006). Proline has been observed to increase thermostability in many mutational studies (van den Burg *et al.*, 1998; Veltman *et al.*, 1996). Lastly, methionine is known as a thermolabile residue because it undergoes oxidation at higher temperatures and its lower frequency has been reported in thermophilic proteins (Kumar *et al.*, 2000; Xu *et al.*, 2003). Hydrophobic group in thermophilic proteins destabilize unfolded forms of the protein and increases with temperature (Britton *et al.*, 1995).

1.7.1.4 Polar and uncharged amino acids

Polar and uncharged amino acids include asparagine, glutamate, cysteine, serine and threonine. They have oxygen, sulphur and nitrogen but they do not ionise or carry an overall charge. Uncharged polar residues specifically serine, threonine and glutamine were found to be much lower in thermophiles (Kumar *et al.*, 2000; Pack & Yoo, 2004). Asparagine and glutamine undergo deamination at higher temperatures (Cantanzano *et al.*, 1997); Serine and threonine are known to interact with water surrounding proteins (Mattos, 2002). The water local protein structure around water binding site is unstable enough to invoke instability at higher temperatures where water interacting with serine and threonine is released. Methionine and cysteine undergo oxidation at higher temperatures and has lower frequency in thermophilic proteins (Kumar *et al.*, 2000; Xu *et al.*, 2003). A strong preference of mutation was observed from mesophiles to thermophiles of Met => Ala; Cys => Ala; Trp => Tyr; Met => Leu; Cys => Val and Cys => Ile (Gromiha *et al.*, 1999). Substitution of methionine to cysteine has proved to be effective in stabilising different amino acids (Estell *et al.*, 1985). A reduction in uncharged polar residues was observed by Sadeghi *et al.*, 2006 to have a specific role in thermostability. Decrease in uncharged polar residues involving asparagine, glutamine, and cysteine that are induced by temperature is likely to minimise deamination, oxidation, and backbone cleavage (Tomazic & Klibanov, 1988).

1.7.1.5 Polar and charged amino acids

Polar charged amino acids include arginine, lysine, histidine, aspartic acid and glutamate. The electrostatic interaction is an important force in maintaining conformational stability in the outer protein part which is influenced by charged amino acids (Kumar *et al.*, 2000).

There are more charged residues found in thermophilic proteins at the expense of uncharged residues (Chakravarty & Varadarajan, 2000; Pack & Yoo, 2004). Increased charged residues retain the hydrogen bonding capability as it provides less labile residues (Sadeghi *et al.*, 2006). The decrease in number of lysine residues which are replaced by arginine and glutamic

acid is a striking feature that stabilizes the exposed structure of thermophilic proteins (Kumar *et al.*, 2000; Chakravarty & Varadarajan, 2000; Das *et al.*, 2006). Arginine has a reduced chemical reactivity and a high tendency to participate in salt bridge interaction (ion pair) which has more stabilizing effect on protein structure. As such, arginine residues are better adapted to higher temperatures than lysine. A higher frequency of arginine and glutamine has been observed in both buried and exposed states of thermophilic proteins. Thermophilic proteins have greater content of charged residues especially at the surface which increase ion interactions and enhances occurrence of salt bridges and ion pairs which are useful in protein stabilisation.

1.7.1.6 Polar and non-polar surface areas

Increased polar surface area has been found to greatly contribute to the stability of thermophilic proteins (Haney *et al.*, 1999). Kumar *et al.* (2000) studied 165 dissimilar monomers. The contributions of polar and non-polar atoms were evaluated between the buried and exposed parts. Distribution of buried and exposed, polar and non-polar were uniform in all 165 monomers between thermophilic and mesophilic proteins which was consistent with findings of Szilagyi and Zavodszky (2000) where no correlation between polarity and temperature was found.

1.7.1.7 Aromatic amino acids

The following are hydrophobic and aromatic amino acids: phenylalanine, tryptophan and tyrosine. Kannan and Visheveshwara (2000) studied 24 structurally similar thermophilic and mesophilic proteins and found a comparatively larger number of aromatic clusters in rigid regions of the surface in thermophilic proteins. Thermophilic proteins in well buried states showed higher frequency of tryptophan (Pack & Yoo, 2004) but occurred with similar proportions in both thermophilic and mesophilic proteins (Kumar *et al.*, 2000).

1.7.1.8 Packing, compactness and hydrophobicity

Russell *et al.* (1997) found better compactness as one of the contributing factors to thermostability after studying citrate synthase isolated from *Antarctic bacterium* in comparison with the same enzyme from a hyperthermophilic host. Compactness is the ratio of accessible surface areas to the surface area of a sphere with the same volume as the protein. The ratio of a better packed protein is smaller. The packing itself involves atoms within subunits and association of subunits with respect to each other. Kumar *et al.* (2000), found no significant difference in packing between thermophilic and mesophilic proteins. But, Scandurra *et al.* (1998) observed an increase in packing density in hyperthermophilic proteins which reduce levels of

thermal changes. A significant difference in number of cavities that correlates to packing and temperature has also been observed in thermophilic proteins (Szilagyi & Zavodszky, 2000). Protein stability was associated with absence of cavities which are capable of accommodating water molecules resulting in increased compactness (Russell *et al.*, 1998).

Hydrophobicity is a dominant driving force in protein folding (Dill, 1990) and thermophilic proteins are known to be hydrophobic (Haney *et al.*, 1997). Hydrophobicity exhibits itself in individual protein chains. Hydrophobicity is computed according to Tsai and Nussinov (1997) as a ratio of buried non-polar surface area out of the total non-polar surface area. Although Kumar *et al.* (2000) found very similar hydrophobicity between thermophilic and mesophilic proteins most studies found hydrophobicity as one of the dominant factors in protein thermal stabilisation.

1.7.1.9 Salt bridges and hydrogen bonds

Several studies have investigated the contribution of salt bridges and hydrogen bonds to thermostability. Hydrogen bonds are divided into three classes which are main chain-main chain (MM Hbonds), main chain-side chain (MS H-bonds) and side chain-side chain (SS H-bonds). Differences have been observed in SS H-bonds and salt bridges in biochemical relevant oligometric states and at their interfaces between thermophilic and mesophilic proteins. An increase in the SS H-bonds and the content of salt bridges was observed at the interface of most thermophilic proteins (Kumar *et al.*, 2000). An increase in hydrogen bonds, ion and salt bridges was also found by Vogt *et al.* (1997) and Scandurra *et al.* (1998). Yip *et al.* (1995) noticed a significant change in the number of salt bridges after studying glutamate dehydrogenase enzyme between thermophilic *Pyrococcus furiosus* and mesophilic *Clostridium symbiosum* with 60 °C differences in melting temperature, and a good sequence and structural similarity. Glutamate dehydrogenase from *Pyrococcus furiosus* and *Clostridium symbiosum* contain 168 and 107 salt bridges respectively representing a 70% increase. Further investigation of salt bridges in glutamate dehydrogenase by Kumar *et al.* (2000a) indicated a higher stabilisation in *Pyrococcus furiosus* as compared to *Clostridium symbiosum*. Szilagyi and Zavodszky (2000) however, found no significant difference in the actual number of hydrogen bonds between mesophilic and thermophilic proteins after studying 64 mesophilic and 29 thermophilic proteins from 25 protein families. They also did not find a significant change in number of hydrogen bonds due to change in temperature. These results agree with those of Russell *et al.* (1998) who found no significant difference in number of hydrogen bonds in thermophilic and mesophilic citrate synthase enzyme from *Antarctic bacterium*.

Russell *et al.* (1997) suggested that deletion or shortening of loops may increase thermostability. However, Kumar *et al.* (2000) did not find correlation between hydrogen bonds, salt bridges, compactness and hydrophobicity against change in number of residues between thermophilic and mesophilic proteins. No correlation was found with change in protein size either due to insertion, deletions or oligomerization. Conflicting observations have also been made on oligometric states of mesophilic and thermophilic proteins.

1.7.1.10 Living temperature

Vogt *et al.* (1997) characterised protein thermostability using the environmental living temperature of organisms. Ponnuswamy *et al.* (1982) found a correlation between melting temperature and both stabilizing and non stabilizing amino acids of proteins after investigating fifteen different proteins. Using a data set of 56 proteins from 16 different families, a relationship was established between stability temperature and polar surface fraction, the number of hydrogen bonds and salt links (Vogt *et al.*, 1997). Based on the same data set, a correlation of 0.91 and a regression equation $T_m = 24.4 + 0.93 T_{env}$ was established between the micro-organism's optimum living temperature in the environment (T_{env}) and its protein melting temperature (T_m) (Gromiha *et al.*, 1999). However correlation was not found by Kumar *et al.* (2000a) between living temperature and structural factors such as oligometric states, chain length, hydrophobicity, compactness, main chain-main chain and side chain-side chain hydrogen bonds and salt bridges. Szilagyi and Zavodszky (2000) also did not observe any correlation between hydrogen bonds and change in optimum living temperature.

1.7.1.11 Ion pairing

Szilagyi and Zavodszky (2000) analysed 16 structural parameters on a data set of 64 mesophilic and 29 thermophilic proteins representing 25 families. Their results revealed highly significant ion pairing in extremely thermophilic proteins; seconded by moderately thermophilic and a weak ion pairings in mesophilic proteins. The increase in the significance of ion pairing from mesophilic to hyperthermophilic proteins was consistent with increase with optimum growth temperature of the organisms. This observation was also confirmed by Vogt *et al.* (1997). Russell *et al.* (1998) further elucidated that localisation into inter subunit ion pair network is a factor in stabilisation.

1.7.1.12 Secondary structures

Vogt *et al.* (1997) analysed secondary structures of 16 families examining structural types such as

helix (H), strand (E), 3_{10} -helix (G), reverse turn (T), β -bridge (B) and coil (C). An overall net increase of H, E and G states and decrease of the T, B and C states was observed as the host living temperature increased. There were more helix or strand states in thermophilic proteins. Szilagyi and Zavodszky (2000) observed an increase in helix content in moderate thermophiles and much higher occurrence of β -sheet content in hyperthermophiles although both showed higher occurrence of α/β type proteins. Kumar *et al.* (2000) also observed higher α -helix content in thermophilic proteins.

This review demonstrates the conflict in findings which studies have come up in attempt to elucidate factors that enhance protein thermostability. There is need therefore for more work to be done in order to settle the controversy that surrounds protein thermostability enhancing factors. These conflicting results are largely due to different methodologies and amount of data that are used in experiments. In addition, factors that enhance protein thermostability are micro-organisms and protein dependent. They differ from one micro-organism to another and within the same organism from protein to protein which brings further complexity in generalising the results. However, general factors could be applied to measure protein thermostability. They would also allow implementation of computational approaches to efficiently determine and predict thermostable proteins. The importance of such results cannot be overemphasised; they would speed up research in application of thermostable enzymes in industry but also ease the burden, labour, cost and resources involved in determining protein thermostability in wet laboratories. The section that follows discusses *in silico* approaches currently employed in determining thermostability.

1.7.2 Approaches for measuring protein thermostability

Prediction of protein thermostability, given a DNA or an RNA sequence is one of the major outstanding challenges in computational biology. Approaches for measuring protein thermostability take advantage of factors that enhance thermostability or their combination to identify thermostable proteins and enzymes. Factors that enhance thermostability discussed in the previous sections include codon usage, amino acid composition; distribution of polar, non-polar, charged, uncharged and aromatic amino acids; packing, compactness, hydrophobicity, salt bridges, hydrogen bonds, ion pairs and living temperature. Results have been conflicting because of different approaches that are used to elucidate these factors. Moreover, the amount of data varies widely and in some cases very insufficient for making general conclusions. Different micro-organisms prefer different factors and even within the same micro-organism factors differ from protein to protein. All this makes the *in silico* prediction of thermostability based on

sequences extremely challenging. To an extent, it is difficult to generalise prediction techniques to all micro-organisms and their proteins because there is no single clear approach for measuring thermostability. Techniques for measuring thermostability include optimum growth temperature (OGT) (Huang *et al.*, 2004) protein melting temperature (T_m) (Kumar *et al.*, 2000) minimum folding energy (MFE) of RNA secondary structures (Zuker & Stiegler, 1981; Markham & Zuker, 2008), codon preference and the (E+K)/(Q+H) ratio (Fariás & Bonato, 2003).

1.7.2.1 Protein melting and optimum living temperature

Thermostability has been described based on melting temperatures which is the temperature at which a protein is active, stable or the protein is at half-life for certain duration (Kumar *et al.*, 2000). Optimum living temperature is the maximum temperature for the environment in which the micro-organism lives. It has been reviewed in detail in section 1.7.1.10. The challenge with this approach is that thermostability of an enzyme varies from one organism to another based on other environmental conditions. The acidity and other chemical compositions of the environment are not the same which have different implications on thermostability of an organism. The environmental temperature itself varies widely making it difficult to establish the exact temperature to work with for a given organism. This is therefore, a very crude way of estimating thermostability as it cannot be generalised to diverse environments.

1.7.2.1.1 Fariás and Bonato ratio

The Fariás and Bonato ratio attempts to exploit the differences in amino acid composition between thermophilic and mesophilic proteins categorised based on optimum growth temperature. After comparing data from 6 hyperthermophiles, 4 thermophiles and 18 mesophiles Fariás and Bonato (2003) observed that there was an increase in glutamate (E) and lysine (K) in hyperthermophiles and a decrease in glutamine (Q) and histidine (H). They found that the ratio of (E+K)/(Q+H) correlated with optimum growth temperature and was above 4.5 in hyperthermophiles, less than 2.5 in thermophiles and between 3.2 and 4.6 in mesophiles. This ratio has been applied in several other studies; for example to predict 31 potential thermostable proteins in *Xylella fastidiosa* which were further classified into 17 groups (van der Linden *et al.*, 2006). Based on this ratio an online tool for measuring protein thermostability was developed and is available for public use (Fariás *et al.*, 2004). The approach is solely based on the sequence composition to predict thermostability. This approach is compromised by the fact that it was deduced based on optimum living temperature whose drawbacks have already been discussed.

1.7.2.2 Minimum Folding Energy

Minimum folding energy is predicted on RNA secondary structures given a DNA or an RNA sequence to elucidate stable functional tertiary structures as shown in Figure 1.7.1. An RNA secondary structure sequence comprises the following bases *A* (adenine), *C* (cystosine), *G* (guanine) and *U* (uracil) where a *T* in the DNA sequence is replaced by a *U* in the RNA sequence. It is a two dimensional structure of bases joined together by interior and exterior edges. Bases are joined by hydrogen bonds (interior edges) to form base pairs. Base pairs are consecutively joined to form a sequence by phosphodiester bonds (exterior edges). Base pairs are mostly formed between *A* and *G* or *A* and *U* which are known as Watson-Crick base pairs while those formed between *G* and *U* are called wobble base pairs. All the base pairs (*AU*, *GU*, *GC*) and their mirrors (*CG*, *UG*, *UA*) are altogether known as canonical base pairs. An RNA secondary structure satisfies the following constraints: (1) it is composed of base pairs (*i,j*) and base pairs must be canonical; (2) there is one to one pairing of bases; (3) bases can pair if they are at least three bases apart ($|i-j| > 3$) and (4) the two pairing bases do not cross. The combination of these rules to join canonical base pairs using interior and exterior edges result in the formation of stackings (helices), hairpins, bulges, interior and multi-branched loops (burification) (Zuker & Stiegler, 1981) as shown in Figure 1.7.2 A and 1.7.2 B respectively. A region bounded on all sides by edges is called a face. A face with single interior edge is known as a hairpin. A stacking region is a face separated by single exterior edges on both sides. However, if it is separated by a single exterior edge on one side and more than one on the other, it is known as a bulge. An interior loop is a face with more than one equal number of exterior edges on both sides.

A structure with more paired bases is more stable as compared to a structure with more unpaired bases such as interior, exterior and hairpin loops. A multi-branched (burification) loop is a face with more than two interior edges. According to basic principles of thermodynamics, the secondary structure formed with the lowest minimum folding energy is believed to be more stable and less affected by external forces (Mohsen *et al.*, 2010).

RNA Sequence

GGGGUAUCGCCAAGCGGUAAAGGCACCGGAUUCUGAUAUCCGGCAUUCGAGGUAUCGAAUCCUCGUAACCCAGCCA

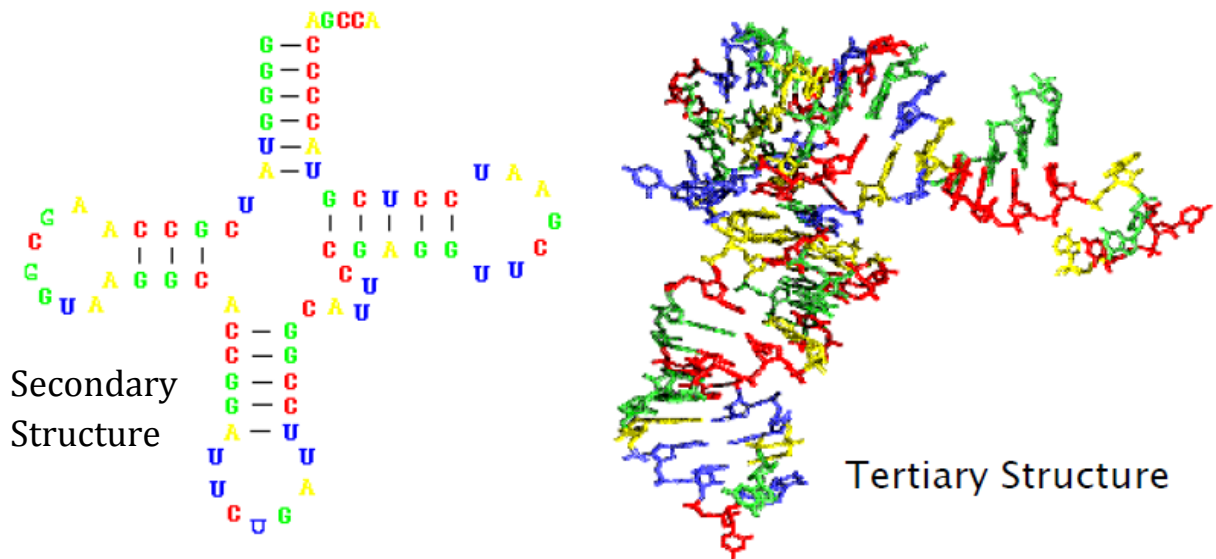


Figure 1.7.1: RNA secondary structure prediction from a sequence (Source: Max Planck Institut Informatik)

Hence, the most stable structure has more negative energy released. Hydrophobic interactions and hydrogen bonding of secondary structures result in formation of tertiary protein structure. Minimum energy prediction however, is not predicted on the tertiary structures because translation forces are weaker in tertiary structures as compared to secondary structures. Their contribution to the formed structure is negligible.

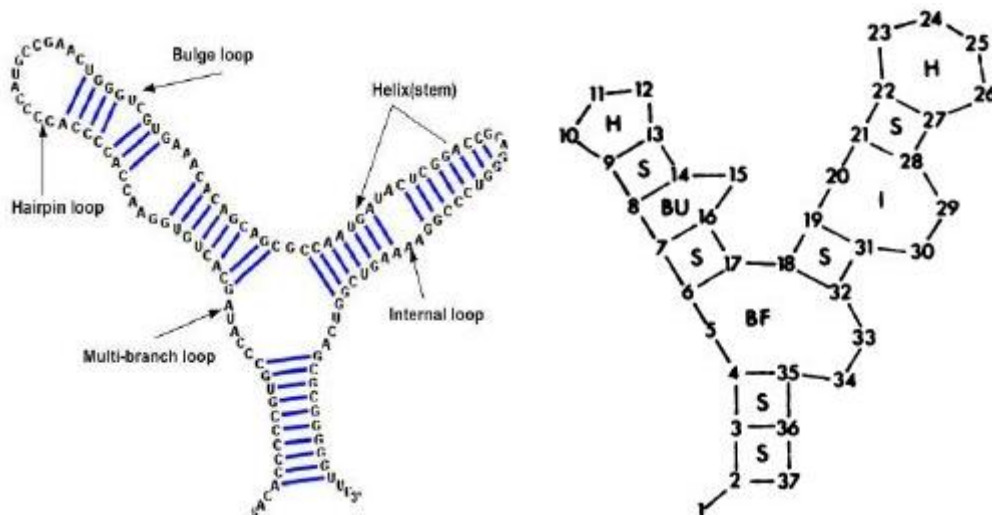


Figure 1.7.2: Components of an RNA secondary structure hairpin (H), interior/internal loop (I), bulge loop (B), multi-branch/burification loop (BU), helix (stacking(S)) (Mohsen *et al.*, 2010; Zuker & Stiegler, 1981)

Prediction algorithms perform thermodynamic optimisation of possible structures to obtain the structure with lowest free energy. Free energy is determined by summing up energy contributions of all base pairs, loops and hairpins. Factors that determine stability of hairpins, bulges and loops

include sequences of loops, nucleotides adjacent to loops, sequences not adjacent to loops and the sizes and shapes of loops (Giese *et al.*, 1998). Empirical results on sequences less than 50 base pairs show that minimum free energy is additive. Individual energy values from the base pairs are determined by melting studies which applied the individual nearest neighbour method whose values have been recently revised and updated (Xia *et al.*, 1998). The free energy of each motif depends on the sequence identity of the most adjacent base pairs. The near-neighbour approach has shown to be sequence dependent.

Minimum folding energy prediction algorithms can be categorised into four groups based on their functionality. The first group predicts hairpins and simple loop formations by performing a basic energy minimisation but excludes multi-branched loops. The second group uses combinatorial approaches to predict all possible secondary structures by piecing them together in all possible ways to find the structure with lowest possible energy. Thirdly, is a group of dynamic algorithms that builds the structure one nucleotide at a time while computing and finding the path with lowest energy values. Dynamic algorithms are so far the most efficient and widely used methods. However, they are extremely expensive in terms of running time and space utilisation. A sequence of length N , has resulting 1.8^N possible structures that can be dynamically computed. Optimised algorithms solve this problem efficiently in $O(N)^3$ in running time and $O(N)^2$ in space (Zuker & Stiegler, 1981). Their limitation however, is that they do not predict energy for pseudo-knots (non nested base pairs). Attempts to predict pseudo-knots dynamically scale up algorithms to $O(N)^6$. The final category is algorithms that apply comparative sequence analysis approaches using multiple sequence alignment. These are more efficient and robust when a number of homologous sequences and structures are available in public databases. However, it is difficult to find such homologous sequences and structures.

Although dynamic algorithms are preferred and widely used, they have several limitations. Two of the limitations arise from the summation of minimum folding energy values from the contributing sequence motifs. The accuracy of these energy values from individual motifs is not clear. There is no empirical evidence to confirm summation prediction for sequences longer than 50 base pairs. The application is based on the assumption that the summation also holds for longer sequences. The third limitation is that near the minimum energy, different foldings are possible and it is difficult to determine the most correct one. Finally, prediction of stable structures is only done on secondary structure because forces that influence folding in tertiary structures are weak and do not significantly influence the final tertiary structure. However, the contribution of the tertiary structure to stability may increase with sequence length. Not all factors that influence folding are fully understood. Even if they were known and fully understood,

taking all possible factors into consideration could further scale up the computational time of algorithms prohibitively.

Several algorithms have been developed for RNA secondary structure prediction. Zuker & Stiegler, (1981) designed the *mfold* algorithm for secondary structure prediction which was later optimised to *UNAFold* algorithm (Markham & Zuker, 2008). Some of the secondary structure prediction programs include *sfold* (Ding *et al.*, 2004), *RNAFold* in the Vienna RNA package (Hofacker, 2003), *MCFold* (Parisen & Major, 2008), *Pfold* (Knudsen & Hein, 2003) and *VSfold* (Dawson *et al.*, 2007). Most of these algorithms are based on the ground breaking algorithm of Zuker and Stiegler (1981).

1.7.2.3 Other thermostability prediction algorithms

Advances have been made in development of computer algorithms to exploit characteristics that differentiate thermophilic from mesophilic proteins. Their complexity range depending on the number of parameters including stringency and crudeness of assumptions made. Most algorithms are specifically applicable to proteins of author's interest although general applicability is increasingly being claimed. Some of the algorithms are hereby described briefly.

Tophan *et al.* (1997) exploited amino acid replacements by translating them into a table of propensity probabilities. In the same way that energy differences are calculated, they calculated propensity differences between wild type proteins and their mutants. They applied the technique on 159 T4 lysozyme mutants on which they achieved an 86% score in attempt to explain thermostability of alpha helices. Carter Jr. *et al.* (2001) related the protein packing to hydrophobicity in which case they calculated correlation between protein packing and unfolding energy change in five proteins: lysozymes, barnase, nuclease, C12 and calbindin. Their findings revealed a 0.8 average correlation on which they could predict the effect of mutations on the hydrophobic core. López de Paz *et al.* (2001) computed the contribution of interaction energies by van der Waals, hydrogen bonds and electrostatics. They used a protein design algorithm PERLA (Protein Engineering Rotamer Library) to create stabilising and destabilising mutations. They constructed a polypeptide structural template by using a custom made side chain. A good level of agreement between predictions and experimental data was found in peptides especially those with beta sheets.

1.8 Genome Rearrangements

Genome rearrangements refer to the permutations of genes or chromosomal sequence order in a genome (Lin *et al.*, 2006). Genome rearrangement is an essential feature for adaptation to

unfavourable conditions such as toxic compound medium and defence from immune system of host organisms. In thermal environments for example, natural transformation is considered to be a survival technique for thermophilic bacteria (Schwarzenlander *et al.*, 2009). Genome rearrangements are responsible for some hereditary diseases due to genomic disorder (Rieseberg, 2001; Stankiewicz & Lupski, 2002). Rearrangements affect gene expression; they may cause complete gene loss if they disrupt open reading frames. In highly naturally competent microorganisms, genome rearrangements are the better measure of vertical inheritance (Darling *et al.*, 2008).

Genome rearrangements are caused by translocations, inversions also known as reversals, duplications, fusion and fission (Pevzner & Tesler, 2003; Hughes, 2000). Inversions and translocation often occur between closely related genomes. Inversions are usually symmetric around the DNA replication axis. When conditions allow, bacteria may return to its initial state or be maintained over evolutionary time scales leading to speciation. What determines where rearrangements occurs is not clear. While Nadeau and Taylor (1984) suggested that rearrangements occur randomly, other observations indicate that they occur in a deterministic manner. Several rearrangement hotspots and clusters have been observed and in some cases reused in independent lineages (Pevzner & Tesler, 2003).

Gene duplication is one major cause of rearrangements. The duplicate copy, normally a homologous sequence on another chromosome, is used as a template to repair the original damaged gene by the bacterial *RecA* protein. Rearrangements occur when recombination is allowed in repeated sequences. Recombination between repeated sequences in the same direction results in duplication. Inversions occur between repeats in the opposite orientation. Deletion of intervening sequences can occur between direct repeats which develop fragments that can potentially be inserted back in the genome at the site of another copy thereby creating a translocation. The *RecA* protein is implicated in homogenization which basically prevents divergence of repeated sequences. It is crucial in maintaining chromosomal integrity for replication but also highly influential in promoting genome rearrangements. Genes may be removed from a genome in one single event in a progressive integration process producing differences between two genomes. Acquisition of new DNA by horizontal gene transfer from other genomes introduces new genetic material and affects the stability of genomes and also triggers other compensating rearrangements. Through horizontal gene transfer, new genes are inserted into the genome thereby disrupting the synteny of genes (Belda *et al.*, 2005). Horizontal gene transfer is known to widely shape the evolution of prokaryotic genomes.

1.8.1 Horizontal Gene Transfer

The diversification and speciation of organisms and survival of resulting species has been explained through the theory of evolution in which among other processes it involves modification of genetic material achieved through mutation, recombination and natural selection. These are reinforced through positive selection which promotes beneficial changes to micro-organisms. Phylogenetic relationships between species are constructed based on these evolutionary molecular markers. The conserved 16S rRNA has proved to be a good choice for constructing phylogenetic relationships between species. However, phylogenetic relationship between organisms based on other orthologous genes result in different, conflicting, and incongruent phylogenetic relationships. Species that morphologically, physiologically or by using other molecular markers are grouped together end up being categorised differently. This eventually has led to the proposal of potential cross species gene transfer through horizontal gene transfer as an alternative explanation of conflicting phylogenetic relationships.

Horizontal gene transfer also known as lateral gene transfer is the acquiring of genetic material from taxonomically unrelated species (Gogarten *et al.*, 2002; Goldenfeld & Woese, 2007). It has brought a new dimension to the study of evolution which predominantly hinged on mutations and natural selection. It has questioned the neo-Darwinian theory of evolutionary synthesis of new traits in micro-organisms alongside the metaphor of the 'tree of life'. The transfer of genes occurs not only among domains such as archaea, bacteria and eukarya, but also across them and in all possible directions (Boto, 2010). As opposed to the slow changes of genetic material caused by continuous mutations, horizontal gene transfer has allowed abrupt large scale change of genetic material and acquisition of new functionality in 'quantum leaps' on the mutational time scale to produce new strains. New ecological, symbiotic, pathogenic functions of genes in the recipient organisms have been observed as a result of horizontally transferred genes. Such new functions are crucial for the survival of the bacteria through adaptation to new or changing environments; or developing resistance to antimicrobial activities such as host immune defence system or drugs. It is generally known through the complexity hypothesis that informational genes (genes involved in transcription, translation and related processes) and house keeping genes are less likely to be transferred. The nature of their functions is unable to fit in a foreign organism. However, operational genes such as those involved in metabolic processes are vulnerable to be horizontally transferred (Jain *et al.*, 1999).

A successful functional horizontal gene transfer is achieved under three conditions: (1) sequence delivery from donor to recipient; (2) incorporation of the new genetic material in the recipient genome and; (3) timely expression of the new genes in required levels in the recipient

(Dutta & Pan, 2002). Horizontal gene transfer occurs through three known mechanisms: conjugation, transduction and natural transformation (Paul, 1999). Natural transformation is the uptake of naked freely available DNA from the environment. It is limited to short DNA fragments. This process is discussed in detail as it occurs in *Thermus* species in subsection 1.8.1.7. DNA transfer through conjugation method requires mediation of conjugal plasmids or transposons. It requires cell to cell contact between distantly related organisms and it transfers longer fragments. Transduction occurs mainly among closely related organisms through phages where cell surfaces of donor and recipient are in contact.

The newly acquired genes through horizontal transfer from a donor form regions known as genomic islands. The GC content and oligonucleotide pattern of genomic islands differs from the global pattern of their recipient genomes. Studies of horizontal gene transfer extend to determine possible donors and time of insertion of genomic islands. For newly inserted genomic islands, their GC and oligonucleotide mean is closer to the global mean of its donor as compared to the recipient genome. However, overtime due to directional mutational pressures and other factors native to the recipient genome such as DNA replication, repair and transcription mechanisms the genomic islands ameliorates to resemble the composition of native genome (Marri & Golding, 2008). Through the process of amelioration, acquired genes adjust to base composition and mutational bias of the recipient genome. The average age of genomic islands is estimating by the degree to which they ameliorated (Lawrence & Ochman, 1997). Several approaches have been used to determine genomic islands. Some of the approaches include: GC content comparison (Groisman *et al.*, 1992) oligonucleotide Markov Chain Analysis (Hayes & Borodovsky, 1998) dinucleotide abundant signature (Karlin & Burge, 1995) codon usage pattern (Karlin *et al.*, 1998) and codon position specific nucleotide composition (Lawrence & Ochman, 1997).

1.8.1.1 GC content

Identification of horizontal gene transfer using GC content analysis largely relies on four factors. First, base composition of bacteria genomes vary widely between species due to biases in directional mutational rates at each of the four bases. Second, genomes that are phylogenetically closer have similar base composition to the extent that they can be grouped accordingly. The third factor is that base composition is uniquely homogenous across the genome (Lawrence & Ochman, 1997). Hence, the GC mean and oligonucleotide usage pattern of a genome is essentially uniform. Lastly, mutational biases have been noted in the base composition of codons mainly at the third position where most changes are synonymous. Hence, if the base composition and codon usage pattern result from mutational biases, horizontally transferred genes will reveal

features foreign to the native genome but typical to the ancestral genome (Lawrence & Ochman, 1997). The GC mean and oligonucleotide usage pattern of a horizontally acquired region significantly differs from the global genomic mean. Depending on the time of insertion, the mean of the genomic islands becomes closer to the ancestral genome than the native genome.

1.8.1.2 Genome and codon signature

Further to this, is the genomic signature defined by relative abundance of dinucleotides computed using odds ratios. The frequencies of dinucleotide XY is given by $P_{xy} = f_{xy}/f_x f_y$ where f_x and f_{xy} are the frequency of nucleotide x and dinucleotide xy respectively. Dinucleotide relative abundance from the same organisms is relatively similar and constant in the bulk of their DNA as compared to those from different organisms. Such uniform structural patterns and genome composition are constrained by factors such as functions of replication and repair machinery, context dependent mutational rates, DNA modifications and base steps conformational tendencies. This uniform influence creates a unique genomic signature. Karlin *et al.* (1998) investigated different dinucleotides between and across prokaryotes and eukaryotes where under-representation and over-representation of certain dinucleotides were observed.

Like genome signatures, codon signature is essentially invariant in large collection of genes. They accommodate amino acid constraints but they are also parallel to genome signatures. The frequency of nucleotides at codon positions 1, 2, 3 is denoted $f_x(1)$, $f_x(2)$, and $f_x(3)$ respectively. Dinucleotides are denoted $f_x(1, 2)$, $f_x(2, 3)$, and $f_x(1, 3)$ and are assessed in a sequence under study through odds ratios $\rho_{XY} = f_{XY} / f_X f_Y$. For codons, this translates to $\rho_{XY}(1,2) = \frac{f_{XY}(1,2)}{f_X(1)f_Y(2)}$; $\rho_{YZ}(1,2) = \frac{f_{YZ}(1,2)}{f_Y(1)f_Z(2)}$ and $\rho_{XZ}(1,2) = \frac{f_{XZ}(1,2)}{f_X(1)f_Z(2)}$.

The average absolute dinucleotide relative abundance difference $\delta(f, g)$ measures the difference between two sequences f, g from the same or different organisms where the sum is based over all nucleotides. A symmetric version is computed for double stranded DNA, denoted ρ_{xy}^* and $\delta^*(f, g)$ for the relative abundance difference. It is calculated over a window to avoid bias of few nucleotides exerting influence on the resulting value by the formula $\delta^*(f, g) = \frac{1}{16} \sum |\delta_{xy}^*(f) - \rho_{xy}^*(g)|$ where f and g are sequences and in some cases f is the contig or fragment. Values for the relative abundance difference are smaller with species as compared across species. G+C content and the p^* -difference for *Helicobacter pylori* are shown in Figure 1.8.1.

1.8.1.3 Codon usage bias

A great variation for codon choices has been observed from organism to organism. Variations in *tRNA* are implicated as a key factor for influencing codon bias in highly expressed genes of *Yeast* and *E. coli* (Karlin *et al.*, 1998a). Foreign genes exhibit deviant codon usage from native genes. This would then be used to classify genes and identify horizontally acquired ones. Codon bias has been assessed within and between gene classes in several ways. It has been assessed in gene expression levels; between genes of different functions; with respect to the size of genes; similarity of amino acid usage or reduced sets of amino acids and for characterisation of alien genes. Codon bias denoted χ^2 detects the presence of codon biasness without necessarily providing the direction of the bias (Lawrence & Ochman, 1997).

Codon bias is measured against the expected values which use all synonymous codons. To show the direction of codon bias, the codon adaptation index is used. The codon adaptation index (CAI) is a quantitative measure for assessing the direction of codon bias usage (Sharp & Li, 1987).

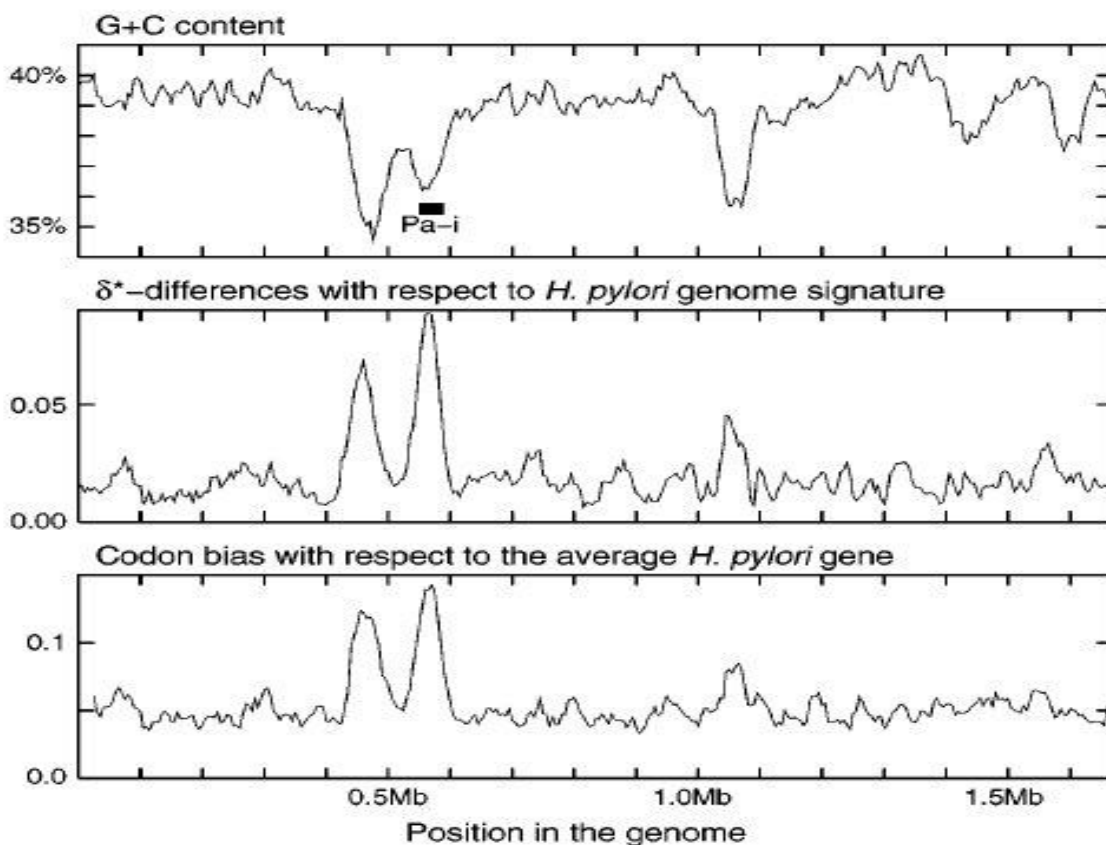


Figure 1.8.1: G+C content, difference, codon bias for *Helicobacter pylori* (Karlin *et al.*, 1998a)

The CAI of a gene of length L is the average $\log \left(\prod_{i=1}^L W_i \right)^{\frac{1}{L}}$ where i is the i^{th} codon and W is the ratio of the frequency of a codon (xyz) to the maximal codon frequency from a reference set

of genes, usually a set of highly expressed genes (Karlin *et al.*, 1998). Figure 1.8.2 is a plot of codon adaptation index against χ^2 of codon usage in *E. coli*, where dots represent native genes and circles are horizontally transferred genes.

In the analysis of 1189 *E. coli* genes with sequences greater than 300 base pairs, there were 1024 native genes represented by dots and 229 horizontally transferred genes, represented by circles. GC content and codon usage are criticised for not being reliable indicators for identification of horizontal gene transfer because they are affected by other factors (Wang, 2001). The GC content and codon usage tend to ameliorate towards the host genome overtime as a result of mutational pressure and DNA repair mechanisms. This affects sensitivity for detecting very old genomic regions and those acquired from relatively closer taxa. Hence, GC content and codon usage need to be reinforced by other mechanisms such as oligonucleotide words occurrence and markov models to improve their sensitivity. Codon usage bias for *Helicobacter pylori* is shown in Figure 1.8.1.

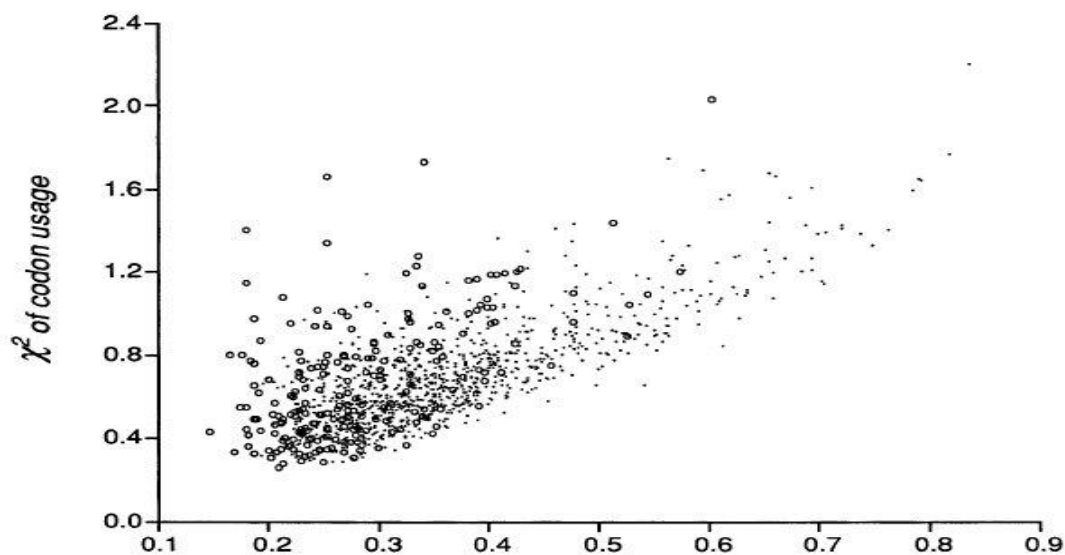


Figure 1.8.2: Codon Adaptation Index against χ^2 for *E. coli* (Lawrence and Ochman, 1997)

1.8.1.4 Occurrence of oligonucleotide words

High or low occurrence and distribution of moderately sized words is used to identify anomalies in the genome. Rare words indicate binding sites for transcriptional control factors or structural defects (Burge *et al.*, 1992). High occurrence often includes parts of repetitive structural, regulatory and transposable elements. In proteins, frequent oligopeptides indicate characteristic motifs shared in some protein families. Important evolutionary tendencies and constraints are depicted when distribution of such words are compared between sequences from different

organisms (Karlin *et al.*, 1996). Oligonucleotide words are used to define genome signatures in an attempt to identify deviant regions (Dufraigne *et al.*, 2005). Oligonucleotide usage pattern has been applied to reveal global features in bacterial chromosomes, plasmids and phages (Reva & Tummler, 2004).

1.8.1.5 Markov Chain analysis

The Markov Chain analysis exploits compositional features of continuous protein coding and non coding regions. Genomic islands are identified by determining sequence regions whose compositional features differ significantly from data set on which they were originally trained (Nicolas *et al.*, 2002; Waack *et al.*, 2006). DNA contigs from the same genome are much similar as compared to those from different genomes. Hence, closely related organisms have similar dinucleotides than distantly related organisms suggesting genome wide factors such as replication and repair machinery, context dependent mutations and DNA modification which pose limits on composition and structural patterns of genome sequences (Karlin *et al.*, 1998a).

1.8.1.6 Horizontal Gene Transfer detection algorithms

Based on the horizontal gene transfer detection criteria discussed above, several algorithms have been developed. These include the SeqWord Genome Browser tool (Ganesan *et al.*, 2008); SeqWord Snifer (Bezuidt *et al.*, 2009); the IslandViewer web portal that combines the prediction results of three algorithms: IslanPick (Langille & Brinkman, 2009), SIGIHMM (Waack *et al.*, 2006) and IslandPath, DIMOB (Hsiao *et al.*, 2003). SeqWord Genome Browser is an online tool for comparative analysis and visualisation of genomic islands based on oligonucleotide pattern statistics. SeqWord Snifer is fully automated and it analyses a large number of genomes for genomic islands at one run based on analysis tetranucleotide compositional biases. SIGIHMM is based on codon usage and Hidden Markov Models to predict genomic islands and their possible donors. IslandPath detects genomic islands using GC content and dinucleotide bias analysis. GC content and dinucleotide bias is calculated on open reading frames (ORFs) or clusters of ORFs respectively.

1.8.1.7 Natural transformation

Out of the three mechanisms for horizontal gene transfer, natural transformation is the most dominant in *Thermus* species. Natural transformation also known as competence is a process by which prokaryotic organisms uptake naked DNA from the environment and incorporate it into their genome to develop new metabolic capabilities. It is considered to be a major survival

technique in extreme temperature environments and has been thoroughly studied in *T. thermophilus* HB27 (Schwarzenlander *et al.*, 2009). *Thermus* genomes have 12 distinct competent genes which code for pre-pilin like proteins similar to type IV pilus biogenesis proteins. Eleven of which were identified and implicated in the transformation machinery in *T. thermophilus* HB27 (Friedrich *et al.*, 2003). DNA binding is achieved by pilQ and is transported through the outer membrane by comEA, pilF and pilA4. DNA is moved through the thick cell wall layers and inner membrane by pilM, pilN, pilO, pilA13 and comEC (Figure 1.8.3) (Schwarzenlander *et al.*, 2009).

Horizontal gene transfer affects the organisation of genes at chromosomal level as new genes are incorporated into the genomes. Introduction of new genes in *Thermus* bacteria due to natural transformation and consequent rearrangements, affect the order of genes and disrupts operon. Organisation of genes on a chromosome is highly linked to their functionality at metabolic level. Co-expressed and functionally related genes in bacteria are grouped into operons or co-localised on the chromosome (Yin *et al.*, 2010).

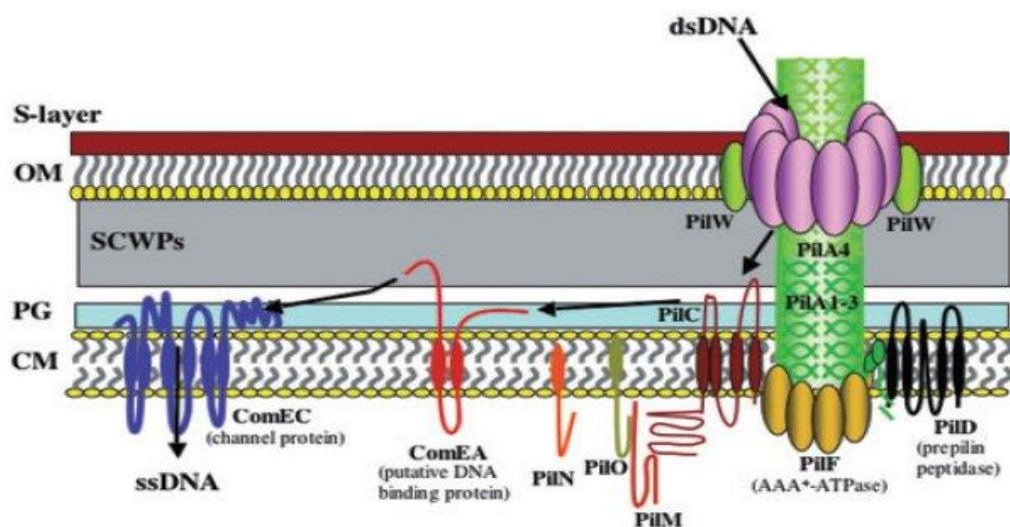


Figure 1.8.3: DNA transport machinery in *Thermus thermophilus* HB27.

Using this machinery naked DNA in the environment is uptaken (Schwarzenlander *et al.*, 2009; Friedrich *et al.*, 2003). It is bound (pilQ), transported through the outer membrane (comEA, pilF and pilA4) and thick cell wall and inner membrane (pilM, pilN, pilO, pilA13 and comEC).

Studies have reported co-localisation and co-expression of genes encoding enzymes catalysing similar metabolic pathways, and that biological networks are built on modules that perform particular functions and are reused in an evolutionary manner (Spirin *et al.*, 2006). Distribution of functionally related genes and breakpoint analysis were applied in this work to measure the level of genome rearrangements.

1.8.2 Breakpoints analysis

Breakpoint analysis was applied to measure the extent to which genes in a genome were reorganised. To achieve this, one genome was identified as a reference genome similar to an out-group in phylogenetic analysis, which acts as a point of comparison for genomes to be analysed. The order and orientation of synteny blocks which are either orthologous genome sequence blocks or genes are compared between the two chromosomes. A breakpoint begins where the order of orthologous synteny or orientation is disrupted. Contrary to the literal interpretation of the 'breakpoint' terminology, a breakpoint is not a particular 'point' where a 'break' as a discontinuation of the genome sequence occurs (Lemaitre *et al.*, 2008). Rather, it is a genomic region between two consecutive synteny blocks of genes or matching genome sequences. The genome rearrangement problem is a combinatorial problem aimed at determining the minimum number of operations made on two genome sequences to make them identical. The rearrangement distance between two genomic sequences is given by the minimum number of deletions, insertions and inversions of genes or genomic sequence blocks to turn the two genomes similar. Associated with the rearrangement distance, is the breakpoint distance, which is half the rearrangement distance because every rearrangement event results in two breakpoints on either sides. Inversion distance defines the minimum number of inversions that needs to be reversed to turn two genomes identical (Watterson *et al.*, 1982).

1.9 Metabolic networks

Biological systems are represented as networks with examples such as protein-protein interaction, gene regulation, and gene co-expression, neural and metabolic networks. Metabolism is a collective term for all biochemical reactions involving substrates and products catalysed by enzymes encoded by genes. The interconnections of biochemical reactions form metabolic pathways which collectively in hierarchical manner create the metabolic network. Topological properties of metabolic networks are essential in understanding the evolution and function of the network. Ancestral organisms have become less dependent on exogenous sources of organic compounds due to the emergence of new metabolic pathways through gene elongation and duplication of genes and operons (Fani & Fondi, 2009). It is through the evolution of metabolic pathways that micro-organisms have developed new sophisticated biochemical processes and metabolic capabilities as new metabolic pathways emerge. Through metabolic evolution, micro-organisms adapt to survive hostile environments such as extreme temperatures, acidity, host immune defence systems and antibacterial chemicals. This draws special attention in this work in attempt to establish dominant evolutionary processes and acquired beneficial metabolic processes and

pathways.

Studies of the evolution of metabolic pathways dates back to 1940s when Horowitz (1945) proposed the retrograde model of pathway evolution in which he theorised that pathway evolves backwards. The theory proposed a chemical environment with enzymes and metabolites, that when a key metabolite is depleted, the organism recruits an enzyme to catalyse another metabolite to replace it. Similarly, if the supply of an enzyme is depleted, another metabolite would be synthesized to produce and replace the depleted enzyme. At genomic level, genes with overlapping specificities such as structural homology or having a common ancestor would cluster into operons. Enzymes for new pathways would be recruited by duplication. However, the theory was limited to explain pathway evolution in a metabolic rich environment only. It also could not account for the development of pathways with metabolites which could not accumulate in the environment long enough (Rison & Thornton, 2002). Hence, Ycas (1974) proposed the patchwork evolution model (Figure 1.9.1) which was later refined by Jensen (1976). This theory proposed that enzymes exhibit broad substrate specificities and catalyse classes of reactions, implying that many metabolic chains synthesising key metabolites may have existed at a low level. Duplication of genes would increase levels of enzymes that would generate more metabolites and their speciation would account for extant pathways. Production of key metabolites from novel intermediates would be allowed from fortuitous evolution of novel chemistry together with biological leakiness. Other pathway evolution theories such as the Granick hypothesis by Granick (1957) and semi enzymatic origin of metabolic pathways (Lazcano & Miller, 1996) were advanced but the retrograde and patchwork models have been the major contenders and a dominant basis of most studies including this study.

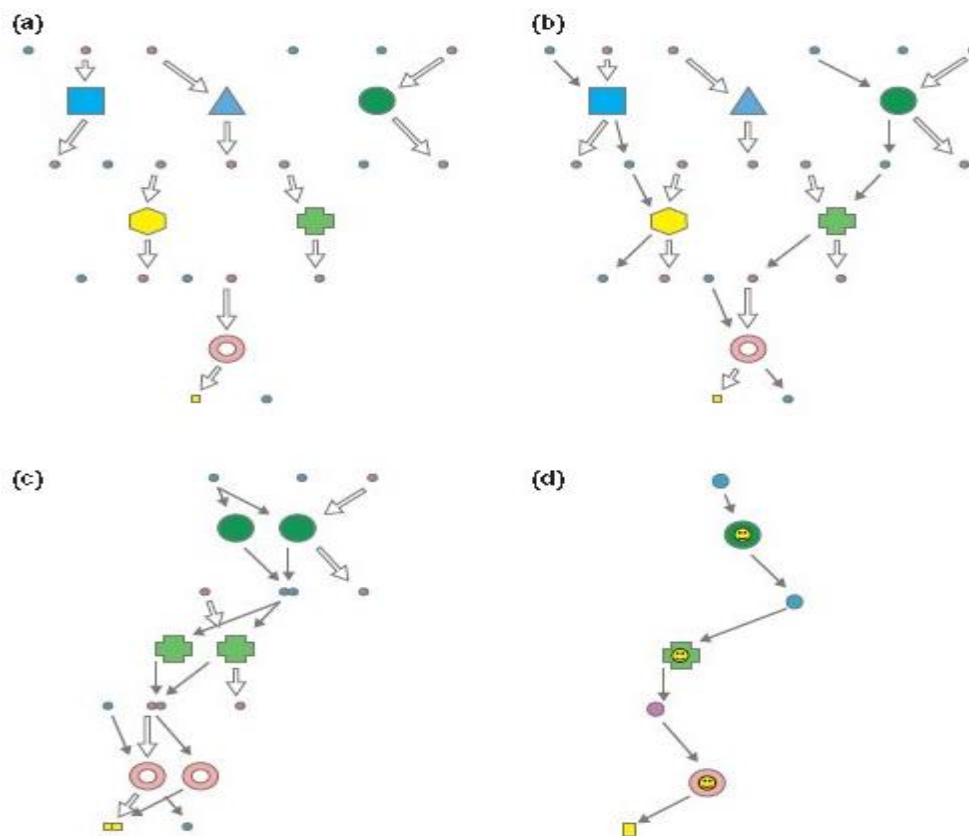


Figure 1.9.1: The Patchwork Model for pathway evolution Rison and Thornton (2002)

Enzymes may have preferred substrates and catalytic mechanisms (a) but are capable of catalysing other reactions (b). Hence many several metabolic reactions may have existed (yellow square, olive circles, green cross and pink doughnut). Duplication of genes (c) paved way for specialisation (d) to account for existing specialised pathways.

To investigate the relationship between structural and functional modules, the metabolic network was integrated using evolutionary associations between genes such as co-occurrence, fused, co-expressed, co-regulated and conserved (Spirin *et al.*, 2006). The resulting metabolic-genome network placed metabolic pathways into an evolutionary and genomic context revealing unknown modules. Analysis showed that evolutionary modules rather than pathways could be considered as regulatory or functional units. Simeonidis *et al.* (2003) analysed metabolic networks using linear programming. A relationship between genome distance, the number of nucleotides separating two genes on the chromosome and enzyme function as characterised by the enzyme commission number was established. No correlation was found between pathway distance - the smallest number of metabolic steps separating two enzymes and enzyme function. However, Rocha *et al.* (2000) demonstrated a significant correlation between the distribution of genes on the chromosome and the physical architecture of cells suggesting a map of cells on the chromosome. In order to investigate evolutionary pressures, metabolic networks were compared across several species such as: archaea versus bacteria; prokaryotes versus eukaryotes; unicellular

versus multi-cellular; motile versus immotile, anaerobic versus aerobic and free living versus host associated (Mazurie *et al.*, 2010). Analysis using graph theory methods and machine learning techniques revealed that organisms with more complex lifestyles are characterised by larger and denser metabolic networks in addition to more efficiently organised cross communications.

1.9.1 Construction of metabolic networks and databases

Efforts to construct metabolic network range from reproduction of simple textbook pathways to complex interactive databases that display pathways, reactions, enzymes, reactants, cofactors and so on (Rison & Thornton, 2002). Genes and enzymes are deduced from genomes and their pattern of interaction is derived from biochemistry. The metabolic network is represented as a connected graph with nodes of substrates and products while edges of enzymes or genes catalysing the substrate. Databases range from individual micro-organism such as EcoCyc and BsubCyc (Karp *et al.*, 2002) to those that generalise pathways from several micro-organisms such as KEGG (Kanehisa & Goto, 2000) and WIT (Overbeek *et al.*, 2000). Kyoto Encyclopaedia of Genes and Genomes (KEGG) and Pathway Tools software are further reviewed and discussed in detail below.

1.9.1.1 KEGG

KEGG links genomic information with high order functional information by computerising current knowledge on cellular processes and standardising gene annotations. Genes in genomes are linked to a set of interacting molecules in the cell such as pathways or high order biological functions. The KEGG project was initiated in May 1995 under the Japanese Ministry of Education, Sports and Culture. Its first release in 2000 consisted of three databases: pathways, genes and ligands. Pathways represent a network of interacting molecules; genes for all completely sequenced genomes and some partial genomes and ligands for collection of chemical compounds, enzymes and enzymatic reactions (Kanehisa & Goto, 2000). These are manually drawn maps for metabolism, genetic information, signal transduction and other cellular processes and diseases. KEGG represent biological systems using line and nested graphs. Nodes in the nested graph can also be graphs. It is used to represent network hierarchy for pathway reconstruction and functional inference. A line graph comprise interchanging nodes and edges to represent coherent complementary metabolic pathways as a network of genes, enzymes or compounds generated by the line graph transformation (Kanehisa *et al.*, 2006).

KEGG has been progressively updated with an increasing amount of genomic and molecular information. The BRITE database was included which is a collection of hierarchies and binary

relations in order to automate functional interpretations of KEGG pathway reconstruction and assist in the discovery of empirical rules involving genome environment interactions. It comprises network hierarchy of orthologs and paralogs, protein families, compounds, compound interactions, drugs, diseases and organisms. The latest version of KEGG comprises 19 integrated databases categorised into system information, genomic information, and chemical information. The system information consists of pathway maps, functional hierarchies and modules, and diseases. Components of genomic information include orthology, organisms, genes for genomes and draft genomes, and sequence similarities. The chemical information has metabolites and other chemical compounds, drugs, glycan, enzymes, reaction, reactant pairs and chemical transformations.

1.9.1.2 Pathway Tools Software

Pathway Tools is software for creating model organisms using integrated information about genes, proteins and metabolic networks known as Pathway/Genome Database (PGDB) (Karp *et al.*, 2002). It was developed by the Artificial Intelligence Centre (AIC) at the Stanford Research Institute (SRI) in the USA as a generalisation of the EcoCyc project to incorporate other genomic data. In order to construct metabolic pathways, it relies on underlying and previously produced databases for *E. coli* (EcoCyc), *Bacillus subtilis* (BsubCyc), general metabolic data (MetaCyc) and other organisms (BioCyc). It performs extensive and complex queries, symbolic computation on these databases to reconstruct new PGDBs based on provided annotation data and genome sequence. This generic approach has several advantages over unique software and database for individual model organisms. Organism specific databases and software are prohibitively expensive to recreate their unique environment for each model organism. Some model organism's software environments use complex algorithm that took years to construct which could therefore be difficult to replicate or could take a long time. Much more, incompatible software would make comparative studies between model organisms extremely difficult.

Components in Pathway Tools version 14 included pathway genome navigators, pathologic program, pathway/genome editors, ontology and the ocelot object database. The genome navigator provides local and web access to query, visualisation and analysis of PGDBs. Pathologic program enables creation of new PGDBs using annotation and sequence information. Pathologic program incorporates several other tools including BLAST that are used to refine metabolic network prediction. Contents of the PGDB can be manually edited to refine its contents, create new metabolic pathways, define transcription factor and binding site interaction or modify protein function using editors. Ontologies define a rich set of classes of attributes and

relationships for modelling biological data. Ocelot object data provides database management system services. Figure 1.9.2 illustrates the functionality of the Pathologic program in Pathway Tools software for prediction of pathways, operons and PGDBs in general. The upper left hand side of the diagram illustrates the input into the program which is basically annotated genomic sequences while the lower left-hand side is the internal information that Pathway Tools uses to predict and construct a new PGDB for the organism in question.

Pathways Tools has been applied in this work to construct PGDBs for *Thermus* species under study in this work. Most analysis focused on operons at genomic level and metabolic pathways. It is of paramount importance therefore, to briefly describe how Pathway Tools predicts pathways and operons. To predict pathways, the software applies enzyme matching between the previously predicted PGDBs which were experimentally verified and the current organism. If it finds matching enzymes that function in a specific pathway, the software assumes that such a pathway also exists in the new PGDB. Subsequent matching enzymes are added to the identified pathway. Enzymes are matched using their names and their enzyme commission (E.C.) numbers. Some pathways remain with missing enzymes with this approach in which case enzyme matching can be manually done using Pathway Hole Filler program. This is a program within pathologic that employs BLAST to match enzymes at sequence level and fill up missing enzymes in pathways. The user can also manually add enzymes to pathways to fill up gaps if all these techniques fail to pick up the right enzyme. Pathologic programme therefore requires complete and correct annotation in order to efficiently predict and construct correct metabolic pathways.

Pathway Tools applied the work of Salgado *et al.* (2000) to predict operons. The prediction approach is illustrated in

Figure 1.9.3. For each contiguous pair of genes, the program examines the direction, whether the pair is within operons (WO) or at transcription unit (TU) boundary (TUB). This is determined by computing the log likelihood intergenic distance (Salgado *et al.*, 2000) whether genes belong to the same function class as matched based on names, E.C. numbers and BLAST. PGDBs provide additional information that is used in operon prediction such as whether both products of genes are in the same metabolic pathways; whether both products of genes monomers are in the same protein complex and if one gene product transports a substrate from a metabolic pathway in which the other gene is an enzyme.

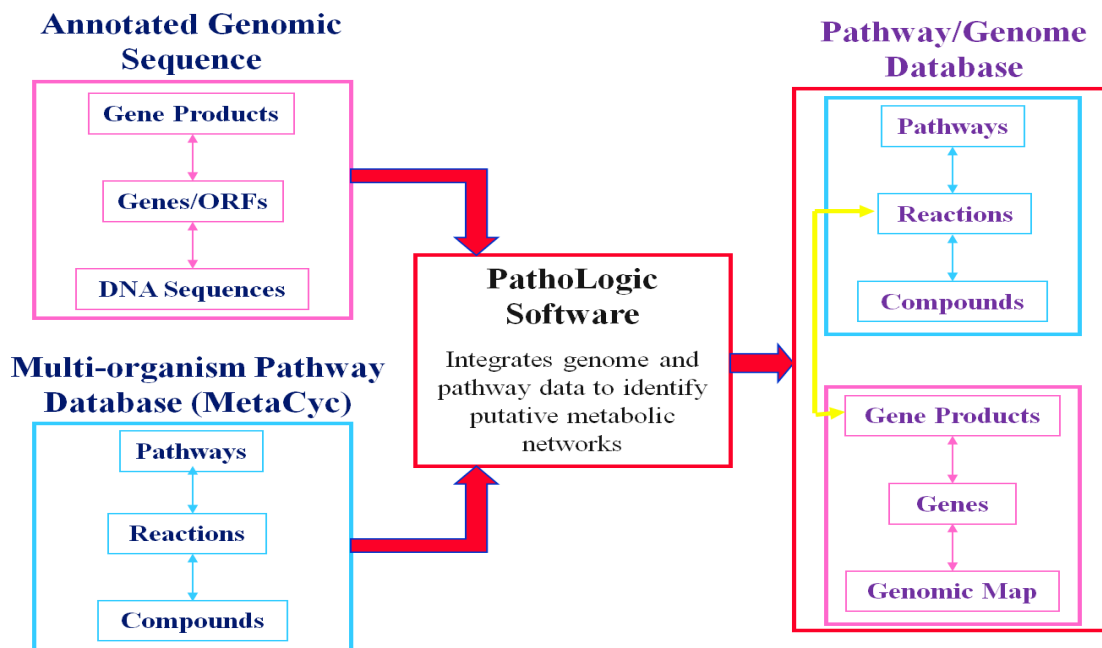


Figure 1.9.2: Software tools pathologic program for constructing PGDBs which includes prediction of pathways and operons (<http://bioinformatics.ai.sri.com/ptools/tutorial/sessions/pathologic/Pathway-prediction-using-PathoLogic/> and the Pathway Tools software user manual)

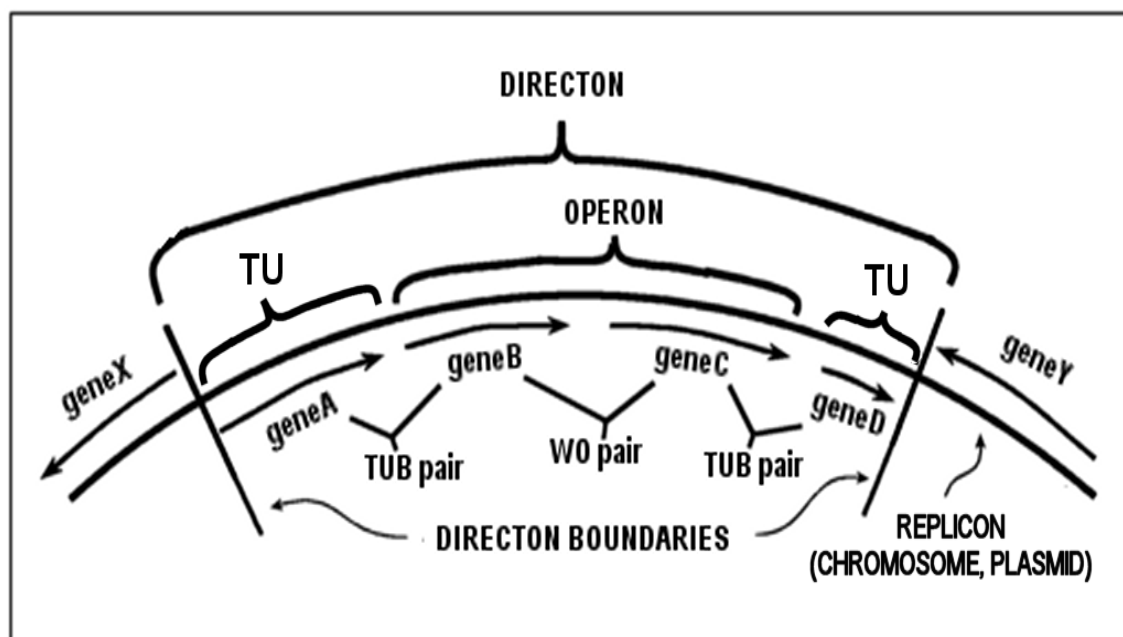


Figure 1.9.3: Illustration of prediction of operons in Pathway Tools (Pathway Tools user manual and <http://bioinformatics.ai.sri.com/ptools/tutorial/sessions/pathologic/Protein-complexes-and-Operon-predictor/>)

The ability of Pathway Tools software to automatically predict and construct pathways gives it an advantage over other static software including KEGG. KEGG pathways for example, are hand drawn which does not provide automated pathway drawing or editing support. It has limited information with no underlying database of flat files that can be used to recreate other model

organisms. However, since KEGG pathways are highly accurate and reliable because they are manually hand drawn. The quality of Pathway Tools software is dependent on the completeness of genome annotation. In this work, Pathway Tools version 14 was used to construct metabolic networks of all analysed organisms. Their networks have been clustered as described in the section that follows.

1.9.1.3 Analysis of metabolic networks

Structural analysis of metabolic networks is central in understanding their evolution. There are a number of ways that are used in measuring the organisation of metabolic networks such as the degree of distribution, small world phenomena, mixing patterns, degree of correlation, clustering coefficient, communities and clusters. The degree of distribution attempts to find the probability distribution that best fits the given data. It is the probability distribution of links per node in a network. The small world phenomenon measures the distribution of shortest paths which indicates the minimum number that can be used to reach any metabolite within the network. Associated to the shortest path are characteristics such as path length which is the average shortest path, the network diameter and the longest shortest path. These are used in analysing the structure and topology of the network. The degree of correlation computes correlation coefficient of degrees of two end points of an edge. It measures the correlation of linkage between nodes of similar degrees. Communities and clusters find size and distribution of connected components which assists in determining the local structure of a network and identifying overrepresented motifs. Such overrepresented motifs have been shown to perform specific tasks such as regulatory control and directing regulatory flux.

1.9.1.4 Metabolic network clustering

Metabolic networks comprise a number of highly connected modules which are linked in hierarchical manner. The coherence of metabolic networks is measured by clustering. The level of clustering is determined by the cross-clustering coefficient also known as the clustering index. It measures the density of triangles in the graph. Having constructed metabolic networks using Pathway Tools, the next step was to compare clustered metabolic networks in order to determine the impact of genome rearrangements. A relationship has been established between metabolic pathways and the conservation of the genomic context (Ogata *et al.*, 2000). Functionally related enzymes are enriched in clusters of chromosomal proximity, co-occurrences and genomic associations. Enzymes of the same metabolic pathway are reported to be grouped into operons and co-expressed. It has been shown that prokaryotic transcription factor genes are co-localised

with their binding sites and essential for formation of regulatory motifs (Kolesov *et al.*, 2007). It has been hypothesised that co-localisation makes transcription faster and more reliable. Significant correlation has also been established between distribution of genes along the chromosome and the physical architecture of the cell mapping on the chromosome (Rocha *et al.*, 2000).

Based on these recent findings, the metabolic network is clustered by determining the ratio of number of enzymes catalysing the same metabolite, within a specific metabolic distance and whether they are co-localised on the genome, co-regulated or co-expressed. Spirin *et al.* (2006) clustered the metabolic network by defining genomic edges and metabolic edges connecting nodes representing genes or enzymes. A metabolic edge exists between two enzymes if they catalyse the same metabolite or are in the same pathway. A genomic edge exists if the neighbours joined by the metabolic edge are co-regulated, co-expressed, fused, co-localised or in the same operon on the chromosome. This is computed for every node and averaged over the entire integrated network to define a cross clustering coefficient as a clustering measure of the network. Simeonidis *et al.* (2003) used correlation between genomic distance and enzyme function to determine clustering of the metabolic network. Genome distance was determined by the number of nucleotides between two genes within a specified metabolic distance, the smallest number of metabolites separating two enzymes. A pathway distance matrix for genome distance categories against metabolic distance was used to study the clustering of functionary related genes. Rison and Thornton (2002) investigated levels of metabolic clustering using pathway distance which is the number of distinct metabolic steps separating two enzymes.

1.10 Summary and aims of the project

This work applied comparative genomics techniques to study *Thermus* bacteria species with focus on *Thermus thermophilus* HB8, *Thermus thermophilus* HB27 and *Thermus scotoductus* SA-01. The bacteria are of special interest because of their ability to live in high temperature environments. Their thermostable enzymes have high potential which can be applicable in biotechnology. Thermophilic enzymes are known to exhibit higher activity and stability and are resistant to denaturing as compared to their mesophilic or synthetic counterparts. This results in higher industrial production and eventual increased profits. The bacteria was also found to reduce heavy metals such as iron and chromium which is useful in eradication of heavy metal pollution, controlling global warming due to reduction in global methane fluxes, reducing toxicity in food among several other uses. High frequencies of natural transformations characterise the bacteria which is one of the mechanisms for horizontal gene transfer that is considered to be a survival

technique in high temperature environments. The acquiring and loss of genes due to natural transformations is believed to affect the arrangement of genes on the chromosome and eventually the coherence of the metabolic network. So far, there have been no studies that investigated factors that enhance protein thermostability, the extent to which *Thermus* genomes have been rearranged due to natural transformation and its impact on the structure of the metabolic network.

There were two major aims of this study namely: (1) to investigate factors that contributed to the speciation of *Thermus* bacteria specifically genome rearrangements, horizontal gene transfer and selective mutagenesis; (2) to analyse factors that enhance protein thermostability in *Thermus* species. The first aim is covered in chapters two and three. Its specific objectives are: (a) to analyse the distribution of functionally related genes on the chromosome; (b) to compute genome rearrangements using breakpoints; (c) to estimate the impact of genome rearrangements on the structure and coherence of the metabolic network and (d) to analyse horizontal gene transfer and its contribution to genome rearrangements.

Distribution of genes, breakpoint analysis and metabolic network clustering are covered in chapter two while horizontal gene transfer is presented in chapter three. Genome rearrangements were investigated by computing breakpoints. The distribution of functionally related genes was analysed by comparing distribution of functionally related genes against the random distribution of genes in a chromosome. Rearrangements were assessed to establish whether they disrupt operons or metabolic pathways. To determine the impact of rearrangements on the metabolic network, metabolic networks were clustered and compared across mesophilic and thermophilic species. Clustering was analysed with respect to living temperature of organisms and genome rearrangements. Several algorithms were used to detect horizontal gene transfer including SeqWord Genome browser and SeqWord Sniffer. Horizontally acquired genomic islands were detected based on the theory that they have an oligonucleotide pattern usage and GC content which are different from their host chromosomes.

In order to study thermostability of *Thermus* bacteria, a suitable measure of protein thermostability for *Thermus* species was determined. This is covered in chapter four and it involved testing several approaches that have been applied on different organisms such environmental optimum growth temperature, protein melting temperature, amino acid preference usage approaches such as Fariás and Bonato and minimum folding energy. The second aim of the study applied optimum growth temperature in estimating thermostability of an organism. This was a crude way of estimating thermostability that could not be applied on individual proteins.

Minimum folding energy of mRNA secondary structures was applied as a measure of thermostability in *Thermus* species. The second aim, its methods and results are presented in

chapter four. The chapter covers: (a) analysis of protein enhancing factors at protein sequence level and (b) investigation of factors that enhance protein thermostability in protein structures. Protein enhancing factors at sequence level involved determination of dominant amino acid residues and their property substitutions between highly thermophilic sequences and less thermophilic orthologous sequences. Analysis of protein structures involved modelling of protein structures based on their orthologous sequences which met the selection criteria. These were orthologs with template structures in PDB with highest sequence identity. Dominant amino acid substitutions and their properties were identified and further investigated in protein structures. The locations and energy contribution of each mutation was assessed in protein structures using FoldX algorithm.









Chapter Two

Genome Rearrangements

2.1 Introduction

Thermus bacteria are characterised by high frequency of natural transformations which is one of the mechanisms for horizontal gene transfer (Schwarzenlander *et al.*, 2009). Through natural transformation, new genetic material is acquired from the environment and incorporated into a bacterial genome resulting into possibly new metabolic abilities useful for survival in extreme thermal environments in addition to other adaptive capabilities. The frequent acquisition of genetic material triggers changes in the order of genome sequence blocks and synteny of genes on the chromosome, a phenomena known as genome rearrangements. Genome rearrangements occur due to deletions, insertions, inversions, transpositions and translocations of genome sequence blocks (Lemaitre *et al.*, 2008). Table 2.1.1 illustrates rearrangement events between the reference and target genomes. The reference genome is used to observe changes in arrangement of genes or genomic blocks in the target genome.

Table 2.1.1: Genome rearrangement events

Reference	Event	Target
A B C 	Translocation	A C B 
A B C 	Inversions	C B A 
A B C 	Deletion	A C 
A C 	Insertion	A B C 

The first column (Reference) shows the order of genes or genomic blocks in a chromosome of a reference genome; the second column (Event) is the name of the rearrangement event. The last column (Target) is the new rearrangement of genes or genomic blocks which occurs in the target genome.

Consider the labelled arrows as genomic blocks or genes in a chromosome. The differences in the order and orientation of the arrows signify the type of rearrangement events which occur. The

reference column in Table 2.1.1 shows the order of genes or genomic blocks in a target genome, the target column represents changes that occur in the target genome and the name of the event is given in the middle column.

Genome rearrangements are classified into intra-chromosomal and inter-chromosomal rearrangements. Inter-chromosomal rearrangements occur between different chromosomes of the same genome or between chromosomes and plasmids, while intra-chromosomal rearrangements occur within the same chromosomes. Rearrangements occur at genome, operon or gene levels. Rearrangements at genomic level involve translocation and inversions of syntenic blocks of the genome sequence spanning over coding and noncoding regions (Pevzner & Tesler, 2003). Operon rearrangements involve splitting of operons into two or more groups of genes which shuffle their order on the chromosome. Rearrangements of genes involve shuffling of genes within operons. This study analysed intra-chromosomal genome rearrangements for the synteny of gene. All bacteria genomes analysed in this study had one chromosome each.

The ordering of genes on the chromosome is not random, but it is highly linked to their metabolic functions. Functionally related genes are grouped into operons and co-localised in proximity on the chromosome (Spirin *et al.*, 2006). Disruption of this order due to acquiring and losing of genes through horizontal gene transfer has implications on the organisation of the metabolic network. The coherence of the structure of the metabolic network defines the extent to which the network evolved. The effect of genome rearrangements on the structure of the metabolic network is therefore investigated to determine its evolution.

The metabolic network is formed from metabolism which comprises all biochemical reactions catalysed by enzymes interconnected to form metabolic pathways. Although the central metabolism comprises similar pathways, the coherence, organisation, modularity and dynamicity of the network structure differ from one organism to another. These metabolic network properties reveal evolutionary pressures experienced by different micro-organisms. Comparison of metabolic networks between different species sheds more light on different evolutionary pressures experienced by those micro-organisms. Distinct evolutionary paths and environmental constraints can be highlighted both qualitatively and quantitatively by differences in metabolic network organisation. Comparative genomic analysis of structures of metabolic networks in archaea, bacteria and eukarya has revealed that different forces and molecular mechanism have shaped them and led to the emergence of new metabolic abilities (Fani & Fondi, 2009).

Several theories have been proposed to explain the evolution of metabolic networks. Two of them are the retrograde and patchwork theories which attempt to explain that the evolution of metabolic pathways is based on gene duplication linking distribution of genes on the

chromosome to the metabolic network structure and functionality (Rison & Thornton, 2002). These theories propose that duplication, elongation and fusion of genes and operons on the chromosome lead to new and specialised metabolic abilities in organisms. Based on these theories, genome rearrangements which affect gene synteny are expected to affect the structure of the metabolic network. As operons are split and genes duplicate, specialised metabolic pathways are formed which affect the coherence of the network structure. It is not clear as to whether the effect of genome rearrangements due to high rates of natural transformations in *Thermus* bacteria contributes to development of new metabolic abilities or whether they occur randomly with no biological implications at all. It is also not clear as to whether genome rearrangements increase with adaptation to thermal environments or at least correlates to thermostability of an organism.

The study of metabolic networks reveals network modularity and existence of network motifs. These among others are useful in characterising functions of genes as applied in functional annotation and prediction of operons, protein functions and new diseases or longevity related genes. In functional annotation, missing genes and parallel modules are identified. This work investigates the evolution of metabolic networks in *Thermus* species and the extent to which metabolic networks have been affected by genome rearrangements.

2.2 Research Questions and Hypotheses

This study investigated the extent to which genes are rearranged in *Thermus* genomes. It further examined whether there is a relationship between genome rearrangements and the level of thermostability as determined by the optimum growth temperature of an organism. The contribution of natural transformations as a major mechanism for horizontal gene transfer in *Thermus* species to genome rearrangements was also investigated.

Genome rearrangements were examined by analysing the distribution of functionally related genes on a chromosome and computing the occurrence of breakpoints. The extents to which functionally related genes are organised define the impact of rearrangements. This has a direct impact on the structure of the metabolic network. Metabolic networks were evaluated to determine different evolutionary pressures that were experienced by organisms and whether they were related to levels of thermostability and rearrangements. The question was whether genome rearrangements caused the emergence or splitting of ancestral and basic biological pathways thereby affecting the coherence of the metabolic network in general. These research questions can be summarised as follows:

1. To what extent have genes been reorganised on chromosomes of *Thermus* genomes? The following specific questions rose in attempt to address this question.

- a. Is there a relationship between genome rearrangements and the level of thermostability?
 - b. Do genome rearrangements disrupt operons and metabolic pathways?
 - c. Does horizontal gene transfer contribute to genome rearrangements in *Thermus* species?
 - d. Are there genomic regions that are more prone to rearrangements? What are the features that characterise genomic regions which are more prone to rearrangements?
2. How does the coherence of metabolic networks in *Thermus* bacteria compare to mesophilic species as a measure of evolutionary pressures?
 - a. Does the coherence of the metabolic network correlate to levels of thermostability?
 - b. Is there any correlation between the coherence of metabolic networks and genome rearrangements?

The first question was addressed by studying the distribution of genes and the analysis of breakpoints. It is discussed in detail in Sections 2.3 and 2.4 respectively. A major assumption that governed the analysis of the distribution of genes was that functionally related genes such as those in the same pathways or acting on the same metabolites are grouped into operons or to some extent co-localised on the chromosome. This was investigated against the assumption that all genes are randomly distributed on the chromosome. The coherence of metabolic networks was assessed by measuring the extent to which the metabolic network was clustered; this is covered in Section 2.5. A cross-clustering coefficient was computed for metabolic network of each organism to compare their coherence. The following research hypotheses guided the analysis of genome rearrangements:

- Genome rearrangements increase with the level of adaptation to higher temperature environments in *Thermus* species.
- Genome rearrangements disrupt operons, pathways and the distribution of functionally related genes on a chromosome.
- Natural transformations as a horizontal gene transfer mechanism affect the arrangement of genes on chromosomes of *Thermus* species.

The second research question on metabolic networks was addressed with the following hypothesis:

- Metabolic networks in *Thermus* species are poorly clustered due to frequent rearrangements.
- There is a correlation between levels of thermostability, genome rearrangements and

coherence of the metabolic network.

Levels of genome rearrangements in *Thermus* species were expected to correlate with the thermostability as previously claimed by Schwarzenlander *et al.* (2009). It was proposed in this study that genome rearrangements enhance thermal adaptation. It was assumed therefore, that genome rearrangements increase with level of thermostability.

The occurrence of breakpoint was analysed to measure levels of genome rearrangements. The expectation was that due to high frequency of natural transformation in thermophilic bacteria, an increased frequency of genome rearrangements would be observed. Functionally related genes in thermophiles would therefore be more shuffled around as compared to mesophiles. Differences in the distribution of functionally related genes between thermophiles and mesophiles would suggest an effect of genome rearrangements which could be associated to their levels of thermostability. Contrary results would suggest that rearrangements occur in a deterministic manner which does not necessarily disrupt metabolic networks but rather strengthen them.

It is not clear as to what triggers rearrangements or in which genomic regions they prefer to occur. Studies have reported that the rearrangements are reused in an evolutionary manner and that certain regions are more prone to rearrangements. Breakpoint positions were analysed to identify possible hotspots and areas prone to their occurrence. Sequences within breakpoint regions were investigated for the occurrence of breakpoints signatures. These are abundantly occurring motifs within breakpoint regions that could be used to characterise breakpoint regions. So far, there have been no studies investigating the levels of genome rearrangements in *Thermus* species partly because of few completely sequenced genomes that were available in public databases. This was therefore, one of the first comparative genomics study that investigated levels of genome rearrangements and their possible impact at both chromosomal and metabolic network level in *Thermus* species.

2.3 Analysis of the Distribution of Genes

The aim of this analysis was to determine whether the distribution of functionally related genes varied with the levels of thermostability. It investigated the differences in distribution of functionally related genes between thermophiles and mesophiles. The distribution of genes was contrasted against the levels of genome rearrangements and thermostability in *Thermus* species.

2.3.1 Gene distribution data set

Analysis of gene distribution was performed on thermophilic *Thermus* bacteria (*Thermus scoto-*

ductus SA-01, *Thermus thermophilus* HB8 and HB27); moderately thermophilic (*Meiothermus silvanus* DSM 9946 and *Meiothermus ruber* DSM 1279); and mesophilic bacteria (*Escherichia coli* K12 and *Bacillus subtilis* str 168). Table 2.3.1 provides more details of the dataset that was used for this analysis. *T. scotoductus* SA-01 was fully sequenced at the University of the Free State in South Africa, which largely motivated this study. All the other sequences were obtained from NCBI GenBank public database.

Table 2.3.1: Data set for gene distribution analysis

No.	Bacteria	Genbank Acc. No.
1	<i>Thermus scotoductus</i> SA-01	NC_014974
2	<i>Thermus thermophilus</i> HB8	NC_006461
3	<i>Thermus thermophilus</i> HB27	NC_005835
4	<i>Meiothermus ruber</i> DSM 1279	NC_013946
5	<i>Meiothermus silvanus</i> DSM 9946	CP002042
6	<i>Escherichia coli</i> K12 substr. MG1655	NC_000913
7	<i>Bacillus subtilis</i> 168	NC_000964

Bacteria organisms with their genome accession numbers which have been analysed for distribution of functionally related genes, rearrangements and metabolic network clustering.

2.3.2 Methods

The distribution of genes was analysed in seven chromosomes for the seven different organisms shown in Table 2.3.1. Distribution of genes on the chromosome was analysed as follows. A hypothetical pair of genes $P_{i,j}$ was selected and the distance between them $d(i,j)$ was measured taking gene direction strands and circularity of bacterial chromosomes into consideration. Consider genes i and j , with gene i starting at position ($start1$) and stopping at position ($stop1$) having hypothetical positions 550 and 2300 on a sequence respectively. These positions are the number of bases between the start and stop positions. The next gene, j has 2500 and 3300 as its hypothetical starting ($start2$) and stopping ($stop2$) positions respectively. Further assume that i and j have the same direction strands. The genomic distance between genes i and j is the absolute difference of the number of nucleotides between the stop of the first gene i and start of the second gene j . Since there could be a possibility that genes may be fused such that the starting position of the next gene j would be smaller than the stopping position of the previous gene i in cases of similar direction strands, the minimum value between all possible calculations of their differences was calculated as shown in formulae (2.3.1) and (2.3.2). Formulae (2.3.3) and (2.3.4) take circularity of the genome into consideration. Hence, the genome distance (d_g) between a pair of genes (i, j) is the minimum number of nucleotides over the four values given that $len(G)$ is the length of genome G (2.3.5).

$$d_1 = |\text{start2} - \text{stop1}| \quad (2.3.1)$$

$$d_2 = |\text{start2} - \text{start1}| \quad (2.3.2)$$

$$d_3 = |\text{len}(G) - (\text{start2} + \text{stop1})| \quad (2.3.3)$$

$$d_4 = |\text{len}(G) - (\text{start} + \text{stop1})| \quad (2.3.4)$$

$$d_g = \min(d_1, d_2, d_3, d_4) \quad (2.3.5)$$

The genomic distances for genes i and j using equation (2.3.1) would be 2200 and 1950 using equation (2.3.2). If the $\text{len}(G)$ is 2,570, 800 nucleotides formulae (2.3.3) and (2.3.4) would give results much greater in which case, the minimum distance of 200 nucleotides would be the genomic distance. If the two genes were in the opposite direction, then the genomic distance would be the minimum value of either the distance between their starting positions or their stopping positions. The computation is shown below in formulae (2.3.6) and (2.3.7). Circularity of the genome was taken into consideration by formulae (2.3.8) and (2.3.9). The minimum distance among all computations provided the genome distance (2.3.10).

$$d_6 = |\text{start2} - \text{start1}| \quad (2.3.6)$$

$$d_7 = |\text{stop2} - \text{stop1}| \quad (2.3.7)$$

$$d_8 = |\text{len}(G) - (\text{start2} + \text{start1})| \quad (2.3.8)$$

$$d_9 = |\text{len}(G) - (\text{stop2} + \text{stop1})| \quad (2.3.9)$$

$$d_g = \min(d_6, d_7, d_8, d_9) \quad (2.3.10)$$

Computation of genomic distance was both central and critical in analysing the distribution of genes. Gene distribution was determined by the formal definition given by equation (2.3.11) and was calculated as follows. Consider a genome G of length $\text{len}(G)$ in nucleotides. Genome distance was divided into six genomic distance categories also known as bins. A bin defined a genome distance range between pairs of genes. A range had start and stop positions denoted x and y respectively. A pair of genes occurring within the genomic distance range x to y was denoted, P^{x-y} . The following distance categories were defined: 0-100; 101-1,000; 1001-10,000; 10,001-100,000; 100,001-1,000,000; 1,000,001-10,000,000. These categories were not chosen randomly but they had biological meanings. The first category (0-100) was approximated by the average distance between genes occurring within the same operon. The second category (101-1000) was determined based on the average length of prokaryotic genes. Ranges in subsequent categories were incremented by an order of the first and second categories (Simeonidis *et al.*, 2003).

Analysis of distribution of genes was aimed at determining differences in the rearrangement of functionally related genes between thermophilic and mesophilic genomes. Hence, the expected distribution was the random distribution of all genes in a chromosome which was regarded as a control and was therefore contrasted against the distribution of functionally related genes. The first step was to determine the expected number of pairs of genes occurring within the same distance categories. Pairs of genes ($P_{i,j}$) throughout the genome were randomly selected w times without necessarily considering their functional relationship. Their distance categories (P^{x-y}) were determined by calculating the number of nucleotides between them as previous shown. Average number of pairs (μ) was calculated by equations (2.3.12). The number of genes to be found within each category could be easily affected by the size of the genome. Hence, genomic distance was normalised by expressing it as the ratio of the genome length $\left(\frac{P^{x-y}}{\text{len}(G)}\right)$. Algorithm 2.3.1 summarises the approach.

1. For $k = 0$ to total number of genes in genome G
2. $i =$ randomly choose gene i between gene at position 0 and the total number of genes
3. $j =$ randomly choose gene j between genes at position 0 and the total number of genes
4. if (i is not equal to j) and (pair not in visitedPair)
5. $d(i,j) =$ compute distance between i and j
6. $P^{x-y} =$ allocate pair i and j to their distance category (increment count of pairs)
7. visitedPairs = keep the visited pair $P(i,j)$
8. Repeat steps 1 to 7 ten times
9. Calculate average values for each distance category

Algorithm 2.3.1: computing expected distribution for all genes in a chromosome

The observed number of functionally related genes was calculated next. These were pairs of genes that coded for enzymes that catalysed similar metabolites or shared similar products within the same pathway. This was a quadratic computation over the total number of genes in the chromosome and it worked as shown in Algorithm 2.3.2.

1. for gene i at position i in chromosome of length $\text{len}(G)-1$
2. for gene j at position $i+1$ in chromosome of length $\text{len}(G)$
3. check if genes i and j are functionally related
4. if genes i and j are functionally related
5. $d(i,j) =$ compute distance between genes i and j
6. $P^{x-y} =$ allocate genes i and j to their distance category
7. Calculate average distance for each category

Algorithm 2.3.2: computing observed distribution of functionally related genes

The algorithm was run once because each run provided the same observed results (μ). The logarithm (β) of observed over expected as shown in (2.3.13) was calculated to show the deviation of observed from the expected distribution of genes.

$$P_{ij}^{x-y} = x > (d(i, j) \leq y); x < y \leq \text{len}(G) \quad (2.3.11)$$

$$\bar{\mu} = \frac{P^{x-y}}{\text{len}(G)} \left(\frac{1}{w} \sum_{n=0}^w \left(\sum_0^{\text{len}(G)} P_{ij}^{x-y} \right) \right) \quad (2.3.12)$$

$$\beta = \ln(\text{observed}/\text{expected}) \quad (2.3.13)$$

2.4 Breakpoints Analysis

Analysis of the distribution of genes generally revealed the extent to which functionally related genes were comparatively shuffled on the chromosome. It examined in particular how co-localisation of functionally related genes was affected by genome rearrangements. However, the analysis of distribution of gene did not provide a precise measurement of levels of rearrangements, particularly the number and positions on a chromosome where breakpoints occurred. Genome rearrangements were determined by the number of breakpoints occurring in a genome specifically translocation and inversions. Precise location of rearrangements events was important in investigating their impact. Identifying precise locations of breakpoint revealed hot-spots and possible breakpoint signatures. Breakpoint signatures could be used to efficiently locate genomic regions which could be most vulnerable to rearrangements.

Breakpoint computation is just one of the ways for determining rearrangements. Genome rearrangements can also be investigated using rearrangement distance. Rearrangement distance is the minimum number of operations (translocation, insertion, deletions and inversions) that are required to be performed to transform one genome sequence into the other. Rearrangement distance and breakpoints are closely related and one can be expressed in terms of the other. Breakpoint distance is twice that of the rearrangement distance. For every event, there are two breakpoints that occur at each end of the sequence. In this work, the number of rearrangement events was determined by using breakpoint analysis on the synteny of genes. For a successful breakpoint analysis, a significant number of genes in the reference genome were required to have their orthologs in the target genome.

Using the breakpoint distance, rearrangements were measured with respect to a reference genome which can be likened to an out-group in the construction of phylogenetic trees. The choice of the reference genome was therefore critical in achieving meaningful breakpoint

analysis. The reference genome was required to be as close to the ancestral genome as possible. Breakpoint analysis results largely depended on the selection of the reference genome. Different reference genomes yield different rearrangement results. This therefore, provides a limitation on the number and types of genomes that can be analysed with a reference genome. In this work for example, it was only possible to analyse rearrangements in *Thermus* species as meaningful results could not be obtained with one reference genome and be generalised to other distantly related species.

2.4.1 Breakpoints dataset

The reference genome sets a limitation on the number and type of organisms that can be analysed for rearrangements. Hence, this analysis was limited to *Thermus* species only. All genomes analysed are available at the NCBI public database under the provided accession numbers. There are over 50 known *Thermus* species; however, at the commencement of this work only three genomes were completely sequenced. The three genomes in which breakpoints were analysed are referred to as target genomes. The reference genome was identified by constructing a phylogenetic tree based on the conserved 16S rRNA sequences of *D. radiodurans* R1, *Thermus* and *Meiothermus* species using Mauve progressive alignment program. *Meiothermus* species were found to be the closest to *Thermus* species and they possibly share a closest ancestor.

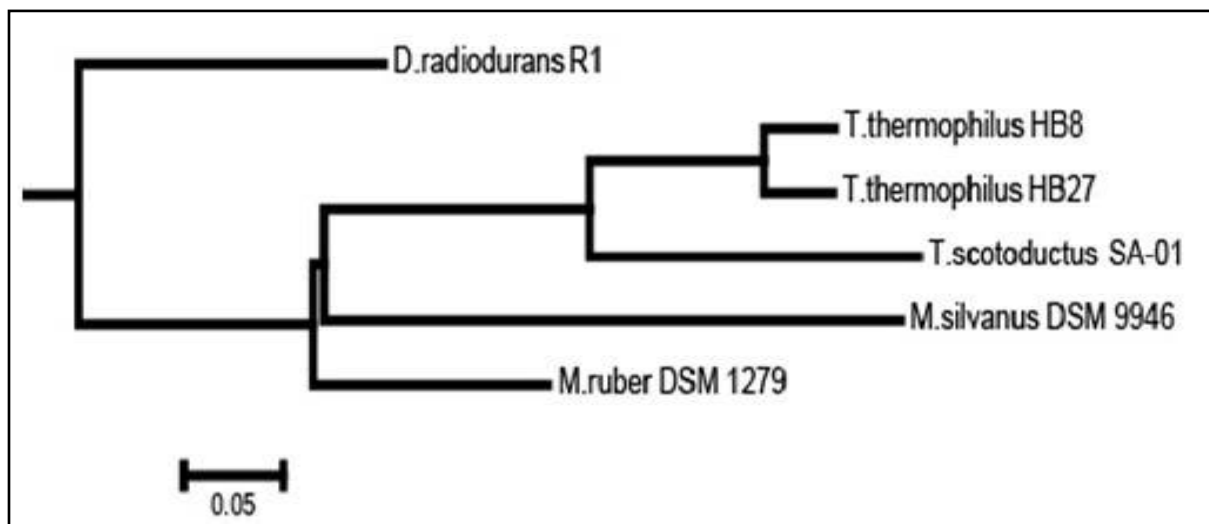


Figure 2.4.1: Identification of *M. silvanus* DSM 9946 as a reference genome for breakpoint analysis (0.05 rates of mutations)

Meiothermus silvanus DSM 9946 was chosen as an ideal reference genome as shown in Figure 2.4.1 for the analysis of breakpoints in *Thermus* species based on its position in the phylogenetic tree. Meaningful breakpoint assessment can only be achieved in closely related organisms

because of the less number of orthologs in distantly related organisms whose similarity is usually relatively weak.

2.4.2 Methods

Figure 2.4.2 illustrates the computation of genome rearrangements. Let G_1 and G_2 be the reference and target genomes respectively and horizontal arrows be hypothetical sequences of genes in a chromosome. Orthologs were calculated for all genes in the reference genome G_1 against all target genomes under study. Gene per gene comparison was performed using Basic Local Alignment Search Tool for protein coding sequences (BLASTp) algorithm implemented in the blastall executable file (Madden, 2002). An assumption was made that since the analysed organisms were closely related; two genes that showed the best BLASTp alignment were orthologs. Two genes were considered to be orthologs if and only if their e-value was less than 0.0001 and the BLASTp score was greater or equal to 100. For simplicity, all genes are shown going in the same direction using horizontal arrows. In reality however, genes in chromosomes do not go in the same direction especially if they are not in the same operon. Let g_1, g_2, g_3 and g_4 be consecutive genes forming a synteny in the reference genome (G_1). These genes may not necessarily be in the same operon.

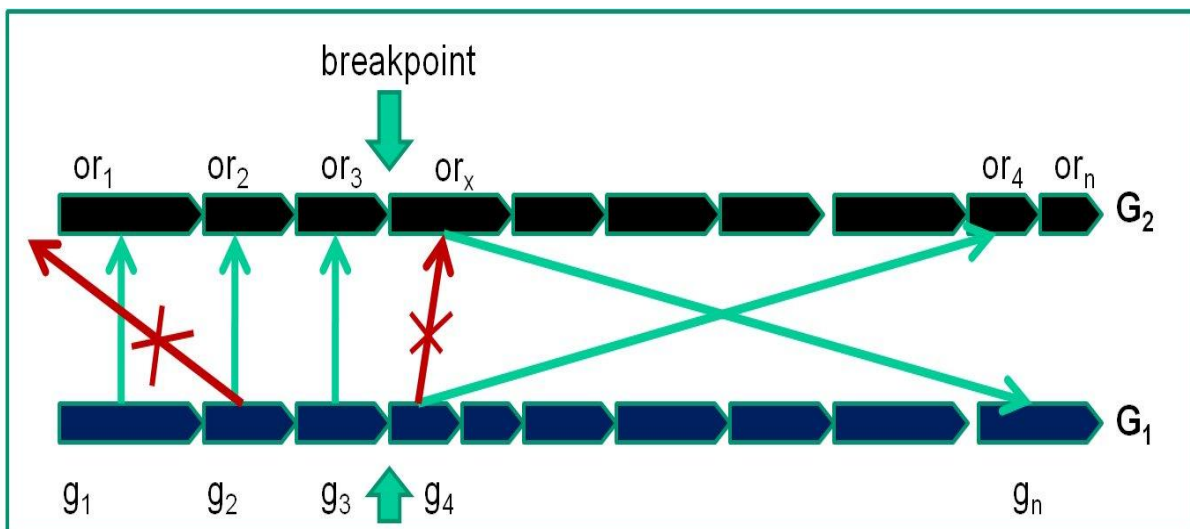


Figure 2.4.2: Computation of breakpoints between the reference and the target genomes.

There is a breakpoint between orthologs or_3 and or_x , if or_x is not an ortholog of gene g_4 and if and only if or_x and g_4 have orthologs elsewhere in G_1 and G_2 respectively. The search begins from g_1 in G_1 (green vertical arrow) and proceed to the next gene g_2 when there was a match or_1 . At g_2 the algorithm checks the adjacent ortholog in the target genome or_2 and the gene before it in case of inversions (red arrow) and so on up to the end of the reference genome.

The breakpoint computation algorithm works as follows. An ortholog (or_1) in target genome G_2 matches the first gene g_1 in the reference genome. The vertical green arrows signify that orthologs in the target genome were found. Then, if the ortholog for g_1 is found, the algorithm moves on to search for an ortholog (or_2) for the next gene g_2 . This ortholog is searched starting from or_1 by checking in both directions, forward and backward as shown by green and red arrows respectively. The red arrows with crosses indicate that there is no match for the next gene in the opposite direction. The two orthologs or_1 and or_2 must be neighbours in the target genome. In this example, the algorithm consecutively finds orthologs up to g_4 . The neighbour to the right hand side of or_3 (or_x) is not an orthologous gene of g_4 . At this position, there is a breakpoint occurring between or_3 and or_x in the target genome on condition that both g_4 and or_x have orthologous relocated elsewhere in the target and reference genomes respectively, otherwise it is either an insertion or deletion. Breakpoints were computed in this manner for the entire chromosome starting with the first gene in the reference genome against all target genomes. Breakpoints were computed for the entire chromosome whilst checking whether they were splitting operons, occurring within genomic islands or pathways. These were compared among *Thermus* species (*Thermus scotoductus* SA-01, *Thermus thermophilus* HB8 and HB27).

Operons and pathways were predicted using Pathway Tools software (Karp *et al.*, 2002). Pathway Tools is software for automatic reconstruction of metabolic pathways. The software predicts metabolic pathways using enzyme commission numbers, annotation information and BLAST search results against other pathway genome databases (PGDB) that were previous constructed and curated from other model organisms such as *E. coli* K-12 (EcoCyc) and *Bacillus subtilis* (BsubCyc). A pair of genes separated by a breakpoint in the target genome was analysed to determine whether the breakpoint occurred in the same operon, genomic island or pathway as illustrated by Figure 2.4.3 and Figure 2.4.4.

Breakpoints were analysed to determine if they were splitting operons. Operons are clusters of genes that are consecutively ordered on a chromosome under the control of one regulatory motif known as a promoter. Operon prediction by Pathway Tools is described in detail in

Figure 1.9.3 in chapter one. Genes within the same operon have the same direction strand. A breakpoint was said to split operons if the genes involved occurred in different operons in the target genome but in the same operon in the reference genome. A breakpoint did not split operons if it occurred between genes in different or same operons in both target and reference genome.

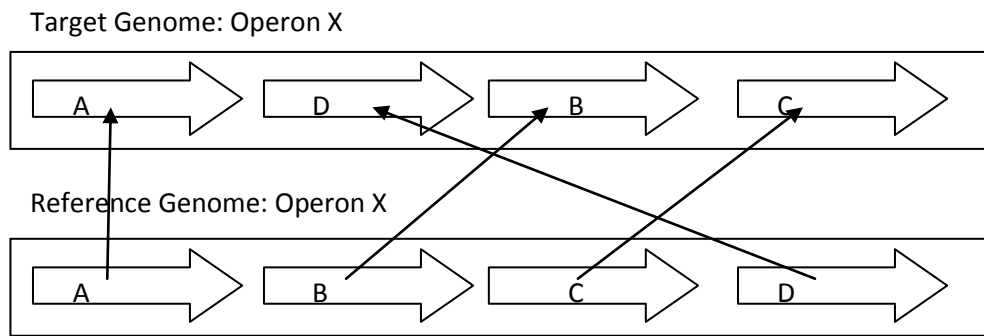


Figure 2.4.3: Breakpoints occurring within the same operon; not splitting operons

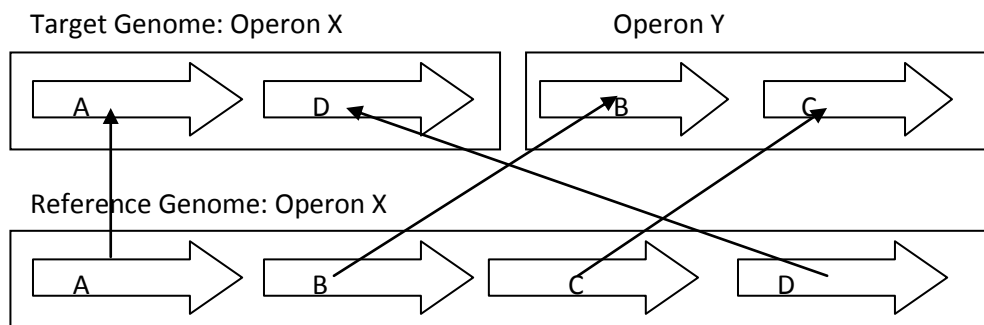


Figure 2.4.4: Breakpoints that splits operons. They occur in the same operon in the reference genomes but different operons in the target genome.

Consider Figure 2.4.3; let the arrows labelled A, B, C and D be genes in operons X in both the target and reference genomes. In this case, breakpoints occur between pairs of genes A and B; and C and D because they are not consecutive in the target genome. However, both these rearrangements are within the same operons; they do not split the operons. However, Figure 2.4.4 presents a different scenario because the breakpoint between A and B does not split an operon in the target genome; it occurs within the same operons. The breakpoint between C and D however, occurs in different operons. It is hard to conclude however, that the breakpoint was responsible for splitting the operon, however the genes are in two different operons X and Y as shown.

These breakpoints were analysed to determine if they occurred in genomic islands. In order to achieve higher detection sensitivity for genomic islands, prediction was done using several algorithms which use different prediction techniques which have different strengths. The SeqWord Genome Browser tool and its semi-automatic realization SeqWord Snifer (Bezuidt *et al.*, 2009) and the SeqWord project website (Ganesan *et al.*, 2008) were applied to detect genomic islands. Positions of breakpoints were plotted on genome diagrams to reveal genome rearrangement hotspot regions. Rearrangement hotspots are regions with high breakpoint occurrence. These regions were analysed to identify possible existence of genomic features that

could be favourable for their occurrence. Breakpoint regions were therefore searched for abundant motifs that could be used as signatures for easy identification of breakpoints.

The aim was to identify possible features that could characterise genomic regions prone to occurrence of breakpoints. A breakpoint region is a noncoding genomic sequence starting from the stop of the first gene (left of the breakpoint) where synteny is broken to the start of the next neighbour gene. Analysis of breakpoint motifs was done from 100 base pairs before the stop of the first ortholog (or_3) to 100 base pairs after the start of the next neighbour gene (or_x). Flanking breakpoint sequences were selected and concatenated for analysis of DNA composition. The analysis of abundant motifs was performed in both coding and noncoding regions of the entire genome and compared to those found in breakpoint regions. Motif frequencies were compared by using MetaLingvo program. WebLogo (Crooks *et al.*, 2004) was used to construct sequence logo to graphically visualise the frequency of motifs

2.5 Metabolic Network Clustering

Analysis of gene distribution and breakpoints provided an indication of changes occurring at genomic level. Metabolic network clustering on the other hand revealed changes occurring at metabolic network levels. These changes were generally responses to the evolutionary pressure that could be related to rearrangements and possibly thermostability of an organism. This section describes an implementation that was used to measure and compare coherence of metabolic networks across bacteria species.

A metabolic pathway is modelled as a graph with edges and vertices representing substrates and enzymes respectively. Although the metabolic network comprises individual pathways, those pathways can be considered separately with distinct properties, the metabolic network itself is essentially one complete network. Hence, the metabolic network was clustered as one complete unit. Analysis of metabolic networks by clustering was motivated by the patchwork evolution model. The assumption was that ancestral organism had enzymes grouped into operons or co-localised on the chromosome which exhibited broad substrate specificity by catalysing different classes of enzymes. This implied that ancestral organisms had a very simple and highly connected metabolic network. However, splitting and duplication of genes and operons resulted into specialisation and development of new metabolic pathways which eventually meant disintegration of the ancestral metabolic network structure.

The level of clustering therefore provided an indication of the extent to which the metabolic network evolved. This work addressed the following research questions. To what extent was the metabolic network clustered? This question attempted to address the closeness to a complete

graph of the metabolic network. Does the level of the metabolic network clustering relate to thermostability? Is there any relationship between metabolic network clustering and the level of genome rearrangements? Poor metabolic network clustering was expected in thermophilic bacteria since their metabolic networks were more likely to have been affected by frequent transformations which were theorised as survival mechanisms in thermal environments. On the contrary, better clustering was expected in mesophilic and moderately thermophilic bacteria. Organisms that went through similar evolutionary pressures were expected to have similar metabolic networks clustering.

The network clustering problem is to determine how close the network topology is to a complete graph. In a complete graph, every node is connected to all other nodes with an edge. The metabolic network was clustered using genome-metabolic associations to create one genome-metabolic network. Two enzymes were associated at genomic level if they were from the same operon or were within a specified genomic distance. Pathway Tools software was used to predict metabolic pathways and operons based on pathway genome databases (PGDB) constructed from other model organisms such as *E. coli* K-12 and *Bacillus subtilis* str 168. Output files for operons and pathways were generated by Pathway Tools software. The files were used as input for a python program to construct a genome-metabolic network which was built by linking genes in the same pathway. Two enzyme encoding genes that acted on the same substrates or had similar products were considered '*functional neighbours*'. In this case, enzymes and genes were represented by nodes while reactions, products and substrates as edges. To simplify the network and avoid creation of unimportant and redundant links, abundant chemicals such as water, ATP, enzyme cofactors and others with more than 10 links between genes were discarded. An edge implied that two enzymes belonged to the same pathway in which case all enzymes in the same pathway formed a complete graph. Genomic edges were those between enzymes in the same pathway with genomic association. A metabolic network would be complete in the unlikely scenario where all pathways comprise enzymes from the same operon. Higher splitting of operons affect the coherence of metabolic networks and is reflected in the quality of clustering. Splitting of numerous pathways could result in poor network clustering. Metabolic networks are believed to start off from such an ideal scenario of completeness. The differences and extents to which the clustering varies indicate levels of evolution for different organisms as they adapt to different environmental conditions.

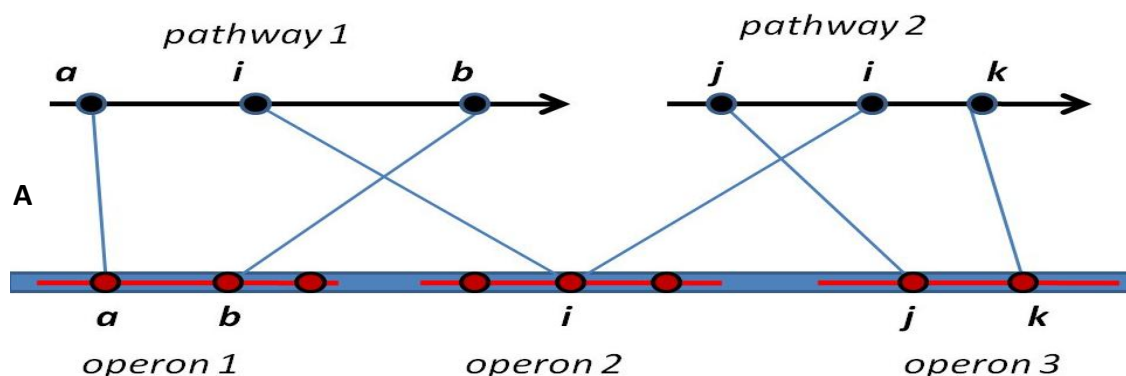
Computation of the cross-clustering coefficient is illustrated in *pathway 1* and *2*, with enzymes *a*, *b*, *j*, *k*, and *i* from different operons in a chromosome as shown in Figure 2.5.1A. Given that there is a metabolic edge from gene *i* to *a*, *b*, *j* and *k* as shown in Figure 2.5.1B, the

cross-clustering coefficient of the node i , (C_i) is the ratio of the total number of possible edges for enzymes in a pathway belonging to the same operon (x) over the total number of edges in the same pathway. Values range from 0 to 1, with 1 being the best clustering for a complete graph representing the most ancestral state of a metabolic network. Nodes j and k ; a and b have genomic edges between them. A metabolic edge exists between enzymes i to a , b , j or k if they catalyse reactions using the same metabolite or they belong to the same pathway. A cross-clustering coefficient (C) given by the formulae 2.5.1 and 2.5.2 is computed by averaging over all nodes h of the integrated network measuring how close the entire metabolic network is to a complete graph.

$$C_i = x/(n(n-1)/2) \quad (2.5.1)$$

$$C = \frac{1}{h} \sum_{i=0}^{i=h} C_i \quad (2.5.2)$$

The computation of the cross-clustering coefficient for purposes of comparison has its limitations especially due to the manner in which operons and pathways are predicted. Metabolic network construction using Pathway Tools software is dependent on the quality of genome annotation and BLAST results from previously constructed pathway genome databases. Poor annotation results in incomplete pathways due to missing enzymes. This is due to several reasons. First, it may be that the organism indeed does not apply the missing enzyme in its pathways. Secondly, it could be that a completely new and novel enzyme may be used for that particular pathway. Lastly, it may also be that the enzyme did not just match its corresponding ortholog. In Pathway Tools software, missing enzymes are handled by a Hole Filler program within the pathologic module. The program uses BLAST to identify homologous genes from other pathways in PGDBs to fill pathways with missing genes.



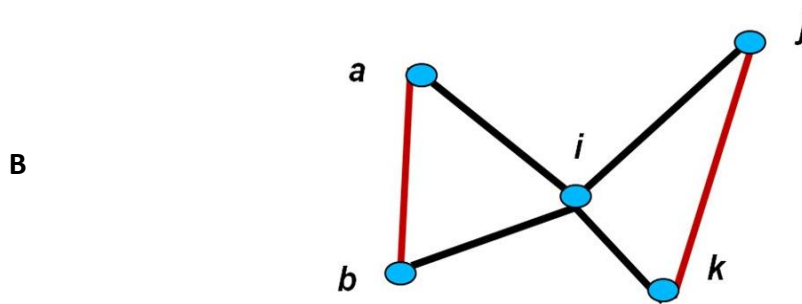


Figure 2.5.1: Genome-metabolic network construction

To address this limitation, only 38 common pathways across all genomes were clustered. The problem of missing enzymes is aggravated with the unbalanced number of pathways that are clustered. By reducing and normalising the number of pathways that are clustered, the number of missing enzymes is also reduced. This standardises analysed pathways and allows clustering of an equal number of pathways to achieve meaningful comparison. The quality of annotation also affects the prediction of operons which also has implication on genomic associations. Hence, the average length of operons was also standardised by estimating an average length of operons to 10, 000 nucleotides.

2.5.1 Data for metabolic network clustering

Metabolic networks were clustered for *E. coli* K-12; *Bacillus subtilis* 168; *Thermus scotoductus* SA-01; *Thermus thermophilus* HB8; *Thermus thermophilus* HB27; *Meiothermus silvanus* DSM 9946 and *M. ruber* DSM 1279. The aim was to compare the level of clustering for mesophilic, moderately thermophilic and thermophilic bacteria in attempt to determine the extent to which the metabolic network evolved. Further to this, was to determine a relationship between the coherence of metabolic networks and levels of thermostability as well as genome rearrangements.

2.6 Gene Distribution Results

Observed and expected distribution of pairs of genes as analysed using genome distance categories are shown in Table 2.6.1 and Table 2.6.2. The values are the total number of pairs of genes within each category. Observed values are pairs of genes within a distance category that shared products or substrates in similar pathways. Observed values were calculated with the assumption that functionally related genes are co-localised on the chromosome. Expected pairs were randomly selected without considering their functional associations. The expected number of pairs of genes was an average from randomly calculated 10 trials. The distance category was normalised over the genome length to compute expected values in order to eliminate the

influence of genome length. An assumption was made that genes were randomly distributed throughout the chromosome.

Table 2.6.1: Observed distribution of pairs of functionally related genes for *T. scotoductus* SA-01 (SA-01), *T. thermophilus* HB8 and HB27, *B. subtilis* 168, *M. silvanus* and *ruber*.

Distance	SA-01	HB8	HB27	<i>B. subtilis</i>	<i>E. coli</i>	<i>M. silvanus</i>	<i>M. ruber</i>
0-100	32	13	45	27	109	27	10
101-1000	1	2	9	2	23	4	1
1001-10000	9	7	68	23	76	7	12
10001-100000	21	16	23	15	27	10	9
100001-1000000	85	35	68	124	182	45	74

Table 2.6.2: Expected distribution of pairs of genes on a chromosome for *T. scotoductus* SA-01 (SA-01), *T. thermophilus* HB8 and HB27, *B. subtilis* 168, *M. silvanus* and *ruber*.

Distance	SA-01	HB8	HB27	<i>B. subtilis</i>	<i>E. coli</i>	<i>M. silvanus</i>	<i>M. ruber</i>
0-100	4.11	3.9	3.80	1.44	0.70	3.20	4.40
101-1000	4.67	4.50	5.60	6.67	1.60	6.90	5.10
1001-10000	57.44	58.50	57.20	57.22	18.30	59.60	57.40
10001-100000	508.78	528	0.2440	0.1192	0.0381	0.1466	483.40
100001- 1000000	1935.0 0	1378.20	1431.60	3718.56	3964.78	2627.2	2463.70

Figure 2.6.1 shows values for the $\ln(\text{observed}/\text{expected})$ distribution of functionally related genes among the seven analysed genomes. This defined the extent to which functionally related genes were observed within each category as compared to the expected number of genes in general. The two smaller bins (0-100) and (101-1000) were combined into a single bin (0-1000). The (0-1000) bin had a higher number of functionally related genes while the largest bin (100,000-1,000,000) had the lowest occurrence of functionally related genes as compared to all other bins for all genomes. Lower values in the largest bin suggested less number of functionally related genes occurring far apart in comparison to functionally related for all genomes.

Gene distribution results showed that in general more functionally related genes were neighbours on the chromosomes in all genomes. There were higher values within (0-1000) category which implied that functionally related genes occurred in proximity. These were genes within the same operons, co-localised or fused but also functionally related. A lower occurrence of functionally related genes was observed within (100,000 - 1,000,000) category in agreement with hypotheses in prior studies.

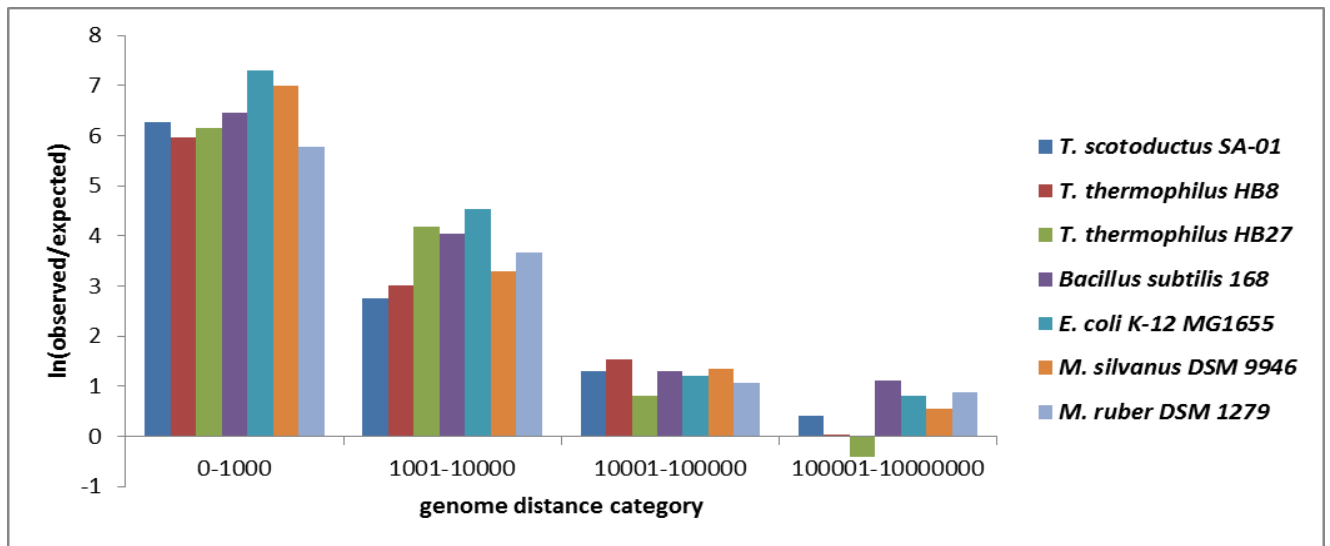


Figure 2.6.1: Distribution of functionally related genes in genomes normalised over expectation

Although the actual values differed within each bin, the distribution of functionally related genes were uniform in mesophilic, moderately thermophilic and thermophilic organisms. The expectation was that there would be a significant difference in the distribution of functionally related genes between thermophilic and mesophilic bacteria due to high levels of natural transformations and possible rearrangements in thermophiles. The earlier assumption based on prior work was that natural transformations would trigger rearrangements of genes on the chromosome. However, genome rearrangements did not appear to have significantly affected rearrangement of functionally genes as they were uniform across genomes.

2.7 Breakpoints Results

Breakpoints analysis results for *T. scotoductus* SA-01, *T. thermophilus* HB8 and HB27 against the reference *M. silvanus* DSM 9946 are shown in Figure 2.7.1. The frequency of rearrangement events were counted in *Thermus* genomes. *M. silvanus* DSM 9946 was selected as a reference genome based on the phylogenetic tree in Figure 2.4.1. These are actual counts of rearrangement events where two genes were neighbours in the reference genome but were not in the target genome.

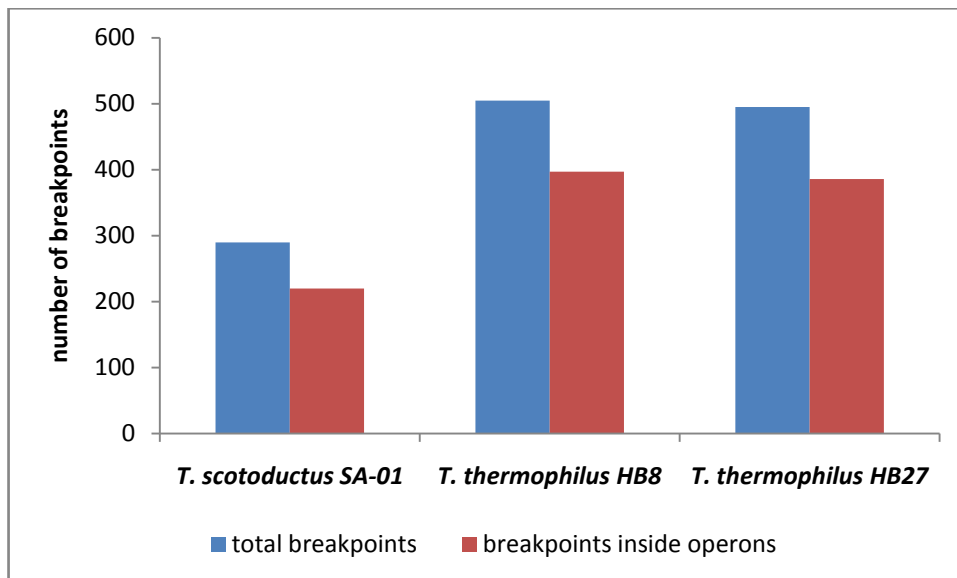


Figure 2.7.1: Frequencies of breakpoint events occurring in *T. scotoductus* SA-01, *T. thermophilus* HB8 and HB27

There were more genome rearrangements, almost double in *T. thermophilus* HB8 (505) and HB27 (495) as compared to those in *T. scotoductus* SA-01 (290). Notably, there were similar frequencies of rearrangements in the two strains, HB8 and HB27 of *T. thermophilus*. The measure of evolutionary changes as observed in Figure 2.4.1 was consistent with the number of rearrangements. The number of genome rearrangements correlated to the environmental optimum growth temperatures for the two *Thermus* species which were between 60 to 65 °C for *T. scotoductus* SA-01; while for *T. thermophilus* HB8 and HB27 temperature ranged from 70 to 85 °C (Balkwill *et al.*, 2004; Opperman & Heerden, 2007; Saiki *et al.*, 1972; Oshima & Imahori, 1974). This observation was consistent with earlier studies which associated high levels of genome rearrangements to levels of thermostability. Most breakpoints occurred within operons in all the three *Thermus* species (Figure 2.7.1) signifying that the distributions of functionally related genes were to some extent preserved.

Further investigations were done to determine whether there was disintegration of operons as an evolutionary mechanism for adaptation. Average operon lengths were computed as presented in Figure 2.7.2. Longer operons were expected in the reference genome as compared to target genome.

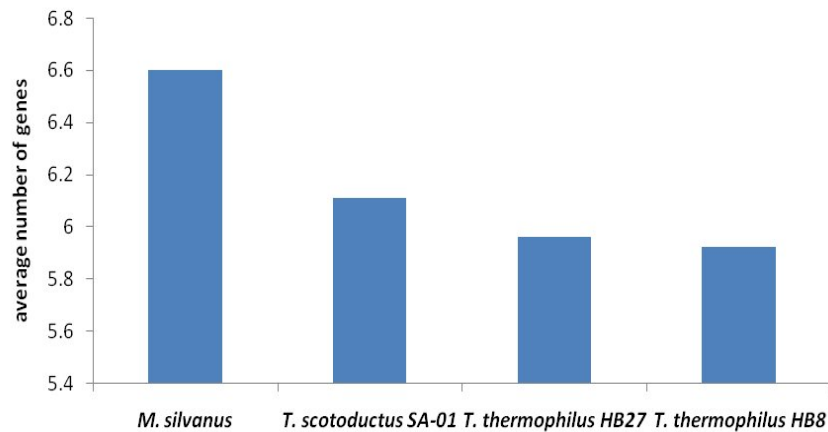


Figure 2.7.2: Average operon sizes calculated in the reference and target genomes

The size of an operon was defined by the total number of genes in it. The average operon size was calculated for all operons in a genome. The average length of operons was observed to be higher in the reference genome, *M. silvanus* DSM 9946 as compared to the three target genomes. However, a statistical test was necessary to confirm statistical significance in differences of operon average length. Operon lengths in the three genomes were considered to be independent samples. Normality test of the samples at 95% level of confidence was conducted using the Shapiro-Wilk test (Shapiro & Wilk, 1965) with the hypothesis that all observations were normally distributed. Normality test results revealed that operon lengths in all three genomes were not normally distributed. Hence, the pair-wise Wilcoxon *t-test*, a nonparametric statistical test was applied as an alternative to the two sample *t-test* that is applied on normally distributed data. The null hypothesis was that the difference in operon length between the reference genome and target genomes was statistically significant. The alternative hypothesis stated that differences in operon length between reference and target genomes were not statistically significant. Results for this analysis are shown in Table 2.7.1. Despite the observed shortening of operons in *Thermus* genomes with frequent rearrangements (Figure 2.7.2), these differences in average operon lengths between the reference genomes and target genomes were not statistically significant. All p-values were greater than 0.05 implying that the null hypothesis was not accepted for all pair-wise considerations.

Genome rearrangements were observed not to significantly affect operon length as per Table 2.7.1. This implied less chance of emergence of new functionality due to specialisation of genes achieved through duplication of genes or operons. This finding is ascribed to the fact that analysed genomes were from the same species. The closeness of these species reduced the chance of divergence and possibility of wide differences in their gene composition and metabolic network organisation. Hence, rearrangements have limitations in deciphering differences among

closely related organisms.

Table 2.7.1: p-values for Wilcoxon rank-sum test for operon length between reference and target genomes

	<i>T. thermophilus</i> HB27	<i>T. thermophilus</i> HB8	<i>T. scotoductus</i> SA-01
<i>T. thermophilus</i> HB8	0.779		
<i>T. scotoductus</i> SA-01	0.532	0.458	
<i>M. silvanus</i> DSM 9946	0.086	0.082	0.306

Although differences in average operon size were not statistically significant among *Thermus* bacteria, it was hypothesised based on average lengths that there were more splitting of operons in *T. thermophilus* as compared to *T. scotoductus* SA-01. Splitting of operons is a well known evolutionary mechanism that enhances specialisation in prokaryotic genes. According to the patchwork evolution model enzymes exhibit broad substrate specificity and catalyses classes of reactions. These are enzymes grouped into operons at chromosomal level which suggest larger operons in ancestral organisms. Splitting and duplication of operons allow enzymes with wide substrate catalytic ability to specialise in specific functions. In this work however, genome rearrangements did not significantly affect operon length.

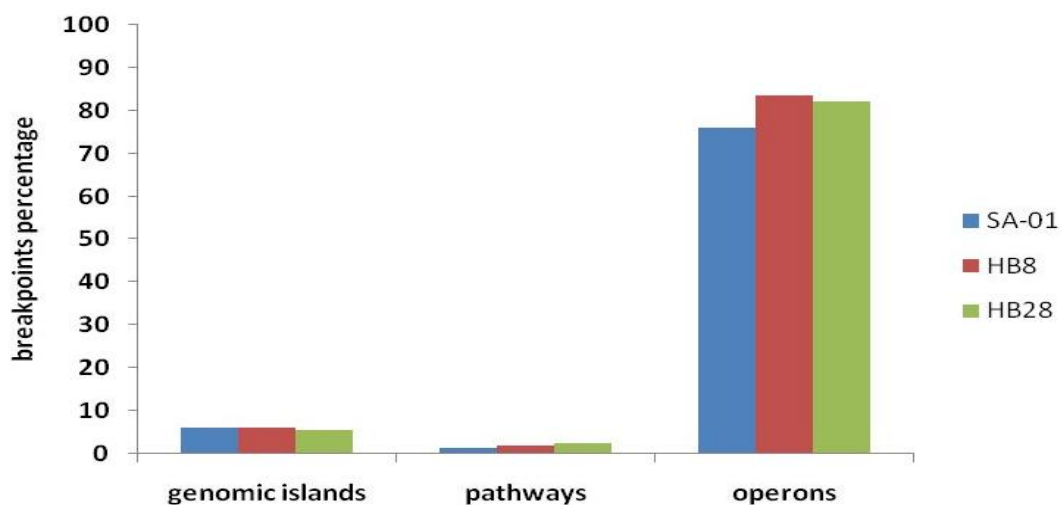


Figure 2.7.3: Breakpoints occurring inside operons, genomic islands and pathways

Further analysis was done to examine whether breakpoints occurred inside operons, metabolic pathways and loci of genomic islands (Figure 2.7.3). A breakpoint was said to occur inside operons if two genes were neighbours within the same operon in the reference genome but their orthologs were not neighbours in the target genome. Over 90% of breakpoints in the three genomes occurred within operon regions. Results revealed that less than 10% of breakpoints

occurred within metabolic pathways. There were more breakpoints within operons than pathways because the proportions of genes that encode for enzymes acting in metabolic pathways are less than genes that encode for generally other functions. In addition, the positions of functionally related genes encoding for enzymes acting in metabolic pathways appear to be highly conserved than genes for other purposes. There were very few breakpoints in genomic islands. The genomic islands are foreign inserts in the host chromosome and it was interesting to investigate the frequency of their rearrangements. If more breakpoints occurred within genomic islands, horizontal gene transfer could be implicated as the cause of genome rearrangements in *Thermus* species. It appeared from this analysis that the mechanisms of frequent genome rearrangements in *Thermus* species were innate and chromosomal specific. Horizontally acquired regions did not contribute to the observed genomic instability and contrary were even more resistant to rearrangements.

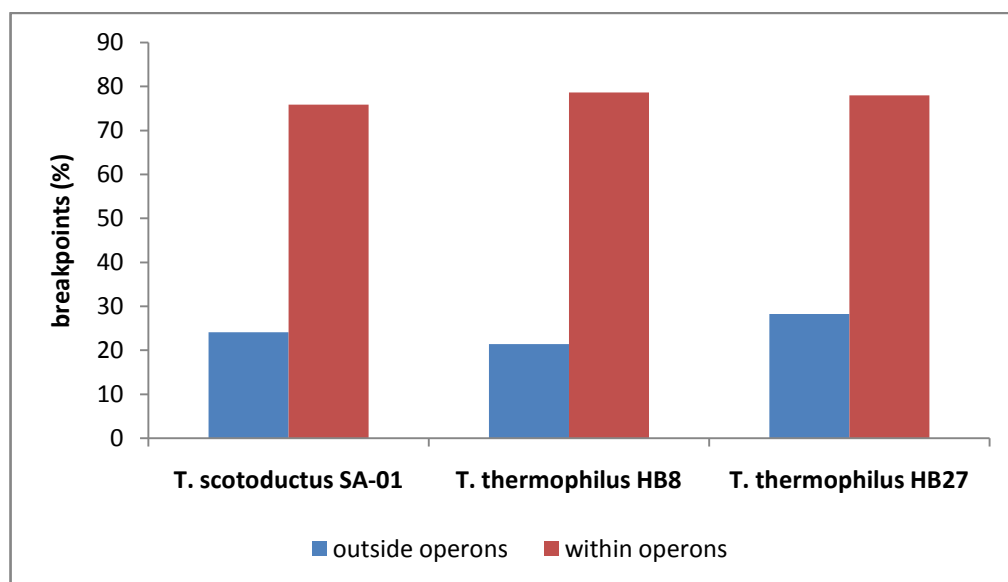


Figure 2.7.4: Breakpoints occurring within operons or externally

This work identified and enumerated the frequency of such events to determine the impact of genome rearrangements in *Thermus* species. Results are shown in Figure 2.7.4 where most rearrangements occurred within operons as compared to between or outside operons in all three *Thermus* genomes. The probability of the synteny of genes being disrupted is much higher within operons than outside operons. Operons may also have relocated within the same chromosome which could preserve the synteny of genes within them but at the same time not significantly contributed to the total number of rearrangements. This finding is consistent with the distribution of genes (Figure 2.6.1) where functionally related genes occurred in close proximity across genomes of different bacteria species.

Inspection of Figure 2.7.5; Figure 2.7.6 and Figure 2.7.7 shows that proportions of regions of horizontally transferred genes were relatively small. As such the impact of horizontal gene transfer on the entire chromosomal rearrangements was essentially insignificant. Genomic islands were identified in regions where the oligonucleotide composition deviated from that of the entire genome. Occurrence of horizontal gene transfer could not be directly associated with genome rearrangements. Rearrangements seemed to occur in a random fashion throughout the chromosomes and that they were rare in genomic islands. Although rearrangements may not be directly associated with the horizontal gene transfer, introduction of new genes due to horizontal gene transfer disrupt gene synteny thereby triggering rearrangements in chromosomes. The same number of genomic islands could have different effect depending on where they occur on the chromosome. The proportion of areas affected by the horizontal gene transfer was relatively small compared to the entire chromosome. The effect of the horizontal gene transfer may not necessarily be in the number and size of the regions, but in locations where they occurred and functions of genes they contain.

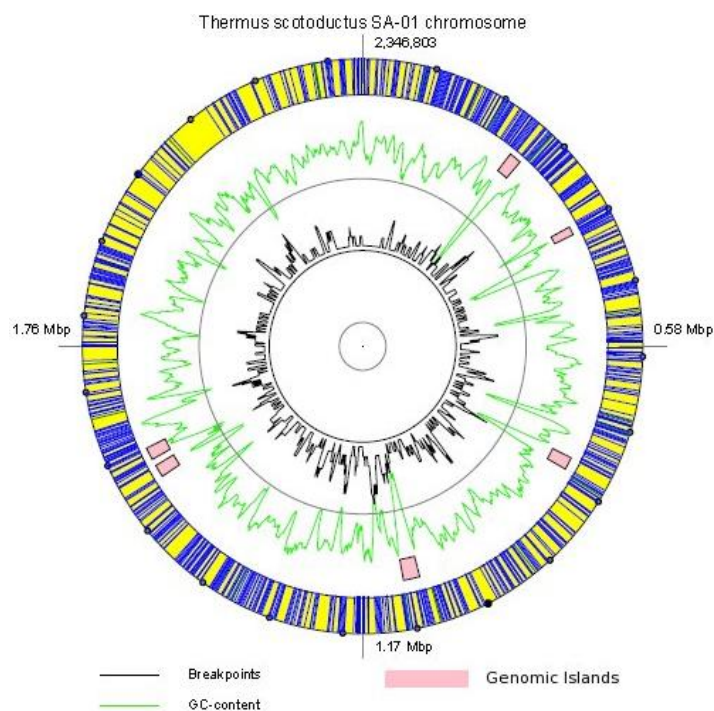


Figure 2.7.5: *T. scotoductus* SA-01 genome rearrangements hotspots with respect to *M. silvanus* DSM 9946

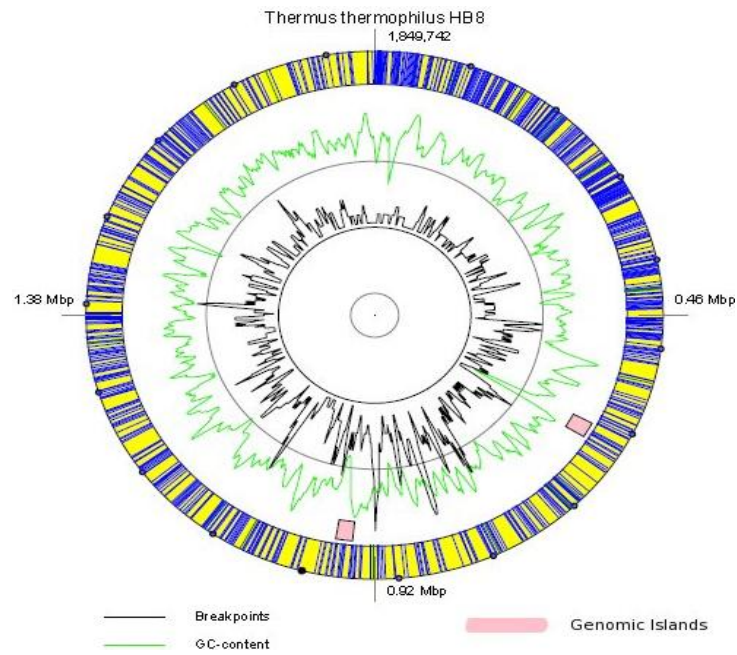


Figure 2.7.6: *T. thermophilus* HB8 breakpoints hotspots with reference to *M. silvanus* DSM 9946

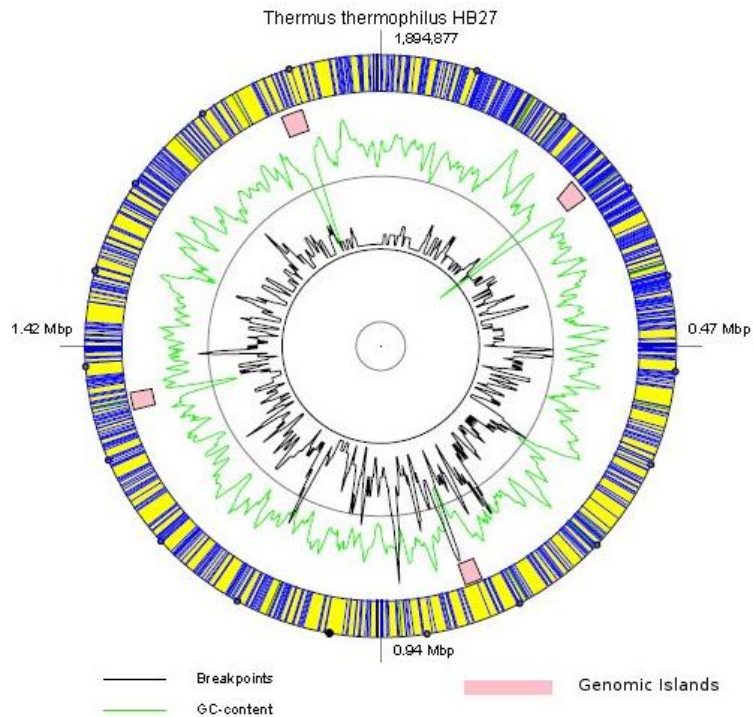


Figure 2.7.7: *T. thermophilus* HB27 breakpoints hotspots with respect to *M. silvanus* DSM 9946

Figure 2.7.5; Figure 2.7.6 and Figure 2.7.7 show breakpoint hotspots, genomic islands and GC content in *T. scotoductus* SA-01, *T. thermophilus* HB8 and HB27 respectively. Although a higher density of breakpoints appears within the first quadrant, they were generally randomly distributed throughout the chromosome. Genomic islands appeared in regions of lower GC content while breakpoints were more frequent in GC-rich loci.

Analysis of frequent motifs in sequences flanking breakpoints is shown in Figure 2.7.8 and Table 2.7.2. An abundant occurrence of the GCGCGC motifs was observed twice as much around breakpoints and in noncoding sequences as compared to the whole genome and coding sequences. The logo represents the conservation with a symbol at each position while the height represents the base frequency in bits.



Figure 2.7.8: The logo of sequence overrepresented in vicinities of genome rearrangement breakpoints

Table 2.7.2: Oligomer frequency in coding and non coding sequences of *Thermus* genomes.

Sequence Region	HB8	HB27	SA-01
Coding	2.1	2.2	1.0
Non coding	4.6	4.0	2.1

Frequencies of GCGCGC oligomer in coding and non coding sequences of three *Thermus* genomes (*T. thermophilus* HB27; *T. thermophilus* HB8 and *T. scotoductus* SA-01) per 10kb window

2.8 Metabolic Network Clustering Results

This section presents results for metabolic network clustering. The methodology was presented in details in the methods section. Common metabolic pathways were clustered based on genomic association which were determined in two ways: occurrence of enzymes within the same operons and secondly, within a defined genomic distance. The genomic distance was defined by the average operon length in nucleotides. The average operon length was estimated at 10,000 bases. This was necessary to eliminate discrepancies that could arise in operon prediction. Common metabolic pathways provided unbiased measure of clustering.

2.8.1 Clustering based on operons and genomic distance

Figure 2.8.1 shows results for clustering of common metabolic pathways in seven genomes based on predicted operons. This eliminated the effect of differences in quality of annotation which was reflected by the imbalanced number of pathways. Thirty-eight pathways which were common in the seven genomes were clustered as shown in Appendix B Table 6.2.1.

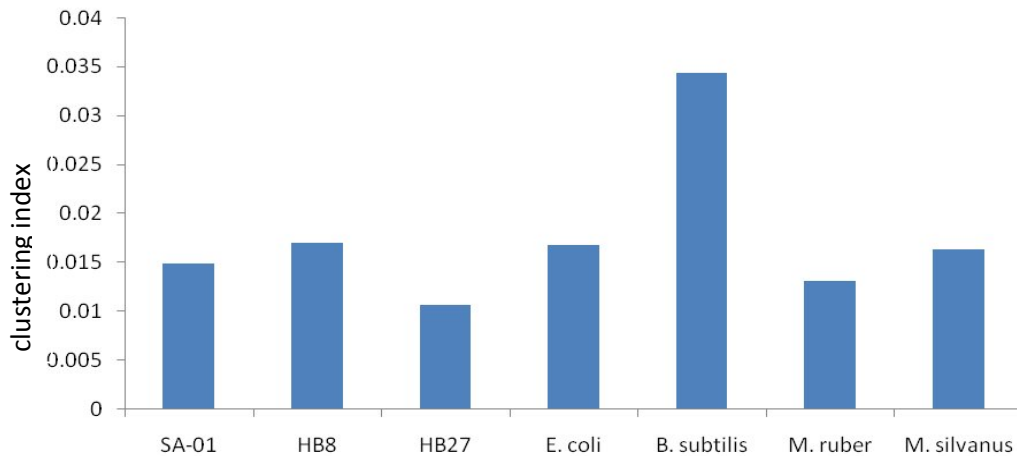


Figure 2.8.1: Clustering of common metabolic pathways using predicted operons

The figure shows clustering based on predicted operons but the difference in clustering is not clear between various bacteria groups. This is due to incomplete annotation of some of the genomes used although to some extent it still reflects differences in clustering, although mesophilic *E. coli* and *B. subtilis* appear better clustered.

Further clustering was done using genomic distance on common pathways as shown in Figure 2.8.2. In this case, instead of using predicted operons as genomic association linkage scores, the average operon size estimated at 10,000 nucleotides was used based on which the clustering coefficient was computed.

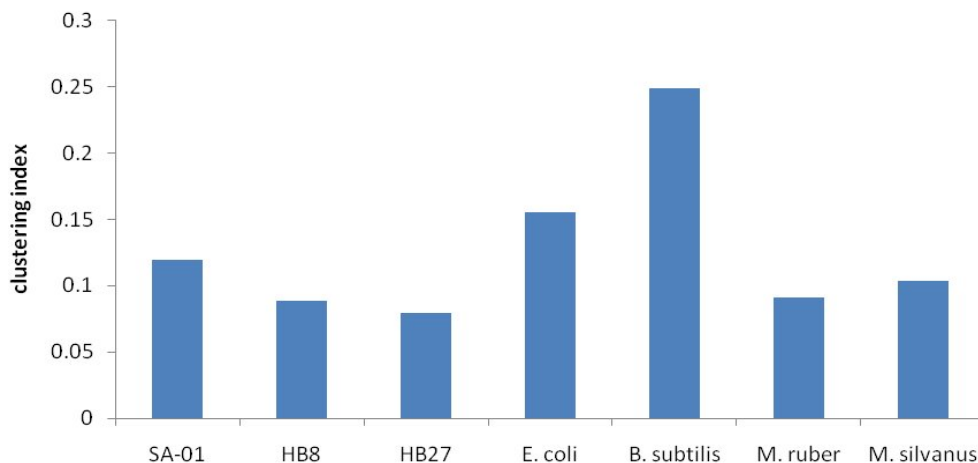


Figure 2.8.2: Clustering of common metabolic pathways using genomic distance

Clustering based on estimated operon size shows differences in clustering between mesophilic *E. coli* and *B. subtilis* as compared to thermophilic and moderately thermophilic bacteria. Poor clustering is observed in *T. thermophilus* HB8 and HB27 as compared to *T. scotoductus* SA-01. There is an overlap in clustering between *Meiothermus* and *Thermus* species as both inhabit environment with similar temperatures.

Differences in clustering were observed, however a statistical analysis was applied to verify the

significance of the observed differences. The working hypothesis was that the pathway clustering indices were independent and normally distributed. Normality test for individual pathway clustering using Shapiro-Wilk showed that the clustering values were not normally distributed for all genomes at 95% confidence level. Hence, the nonparametric pair wise Wilcoxon *t-test* was applied to determine whether differences in levels of clustering were statistically significant at 95% level of confidence. Table 2.8.1 and Table 2.8.2 are *p*-values for clustering based on predicted operons and genomic distance association respectively.

Table 2.8.1: Pair wise Wilcoxon *t-test*, *p*-values results for common pathways clustering based on predicted operons

	<i>B. subtilis</i>	<i>E. coli</i>	HB27	HB8	<i>M. ruber</i>	SA-01
<i>E. coli</i>	0.0089					
HB27	0.0269	0.7448				
HB8	0.1377	0.8181	0.9284			
<i>M. ruber</i>	0.0269	0.8287	0.8368	0.8287		
SA-01	0.1181	0.4851	0.8287	0.8368	0.8181	
<i>M. silvanus</i>	0.1181	0.5290	0.8287	0.9281	0.8181	0.9184

Clustering of metabolic pathways based on predicted operons and average operon distance estimated above showed a better clustering in mesophilic bacteria (*E. coli* K-12 and *B. subtilis* str. 168) as compared to the thermophilic bacteria (*T. scotoductus* SA-01, *T. thermophilus* HB8 and *T. thermophilus* HB27) and moderately thermophilic (*M. silvanus* DSM 9946 and *M. ruber* DSM 1276). Although there were differences in the actual values between the two types of analysis, metabolic pathways for thermophilic and moderately thermophilic species were relatively poorly clustered in both cases.

Table 2.8.2: Pair wise Wilcoxon *t-test*, *p*-values results for common pathways clustering based on genomic distance association

	<i>B. subtilis</i>	<i>E. coli</i>	HB27	HB8	<i>M. ruber</i>	SA-01
<i>E. coli</i>	0.0871					
HB27	0.0059	0.4806				
HB8	0.0059	0.0706	0.4806			
<i>M. ruber</i>	0.0104	0.3973	1.0000	0.5667		
SA-01	0.0349	0.8214	0.5663	0.2137	0.5602	
<i>M. silvanus</i>	0.0349	0.7902	0.5663	0.2975	0.6420	0.9889

Levels of clustering were consistent with the thermophilic lifestyle as determined by environmental optimum growth temperature of the organisms in Figure 2.8.2. Comparison of levels of clustering among *Thermus* species showed poor clustering in *T. thermophilus* HB8 and *T. thermophilus* HB27 as compare to *T. scotoductus* SA-01. *M. silvanus* DSM 9946 and *M. ruber* DSM 1279 showed similar clustering to that in *T. scotoductus* SA-01. This complied with their optimal growth temperature since *T. thermophilus* HB8 and *T. thermophilus* HB27 were isolated

in comparatively higher temperature environments. The level of clustering suggested the extent to which the metabolic network evolved. Metabolic networks for *Thermus* species therefore, seemed to have undergone more evolutionary changes in their organisation which was possibly due to shuffling and disintegration of pathways to suite their thermophilic lifestyle.

2.9 Discussion

Analysis of gene distribution showed that functionally related genes occurred in proximity in chromosomes of all genomes despite differences in occurrence of genome rearrangements and levels of thermostability. There were relatively more pairs of functionally related genes occurring within the 0-1000 nucleotides distance category. However, there were less functionally related genes in genome distance categories with distance greater than 100,000 nucleotides. This agreed with the earlier hypothesis as per prior studies which found chromosomal proximity, grouping into operons and co-expression of functionally related genes (Spirin *et al.*, 2006). Comparison among mesophilic, moderately thermophilic and extremely thermophilic bacteria yielded a uniform distribution of functionally related genes over all genome distance categories. This showed that synteny of functionally related genes were not strongly affected by the frequency of genome rearrangements in thermophiles, probably due to evolutionary counter-selection of advert rearrangements. All genomes were found to have more functionally related genes within the 0-1000 distance category. Insignificant differences were however observed in proportions of functionally related genes within the 0-1000 category. *E. coli* K-12 and *M. silvanus* DSM 9946 had the largest number of functionally related genes while *T. thermophilus* HB8 had the lowest.

Comparison of numbers of functionally related genes within 0-1000 among *Thermus* species revealed that *T. scotoductus* SA-01 had slightly higher level of clustering of functionally related genes than *T. thermophilus* HB8 and *T. thermophilus* HB27. This difference can be explained also by other evolutionary factors. *Thermus/Meiothermus* genomes comprise chromosomes, megaplasmids and small plasmids. The number of plasmids per genome differs between strains. *T. scotoductus* SA-01 comprises one chromosome and a plasmid TSCp8. *T. thermophilus* HB8 and HB27 possess additional large plasmids. However, the chromosomes of the latter organisms are shorter than that of *T. scotoductus* SA-01. There is a shorter plasmid in *T. thermophilus* HB8 that is similar to TSCp8, but do not have the same genes (Gounder *et al.*, 2011). For the majority of genes present in *T. scotoductus* SA-01 their orthologous counterparts were found in *T. thermophilus* chromosomes and their plasmids. Functional analysis of genes located on the megaplasmids of *T. thermophilus* showed that they coded for several metabolic pathways, namely: coenzyme B12 synthesis and metabolism; adenosylcobalamin biosynthesis and

adenosylcobalamin salvage pathways; dATP, dGTP and dUTP biosynthetic pathways; neurosporene and siroheme biosynthesis. Other genes encoded different metabolic enzymes: acylCoA dehydrogenases, isomerases, oxidoreductase, glucosidases, galactosidases and several others. All these genes are spread on the chromosome of *T. scotoductus* SA-01 that probably was the case with the common ancestor of *Thermus* species. This may explain why *T. scotoductus* SA-01 has a slightly higher number of clustered functionally related genes as compared to *T. thermophilus* HB8 and HB27 (Figure 2.8.2).

Breakpoint analysis revealed fewer rearrangements in *T. scotoductus* SA-01 as compared to *T. thermophilus* HB8 and *T. thermophilus* HB27 with reference to *Meiothermus silvanus* DSM 9946. This was strengthened by the fact that fewer genes moved from *T. scotoductus* SA-01 chromosome to its plasmid which is smaller than plasmids in *T. thermophilus* HB8 and *T. thermophilus* HB27. The reference genome obtained from computing a phylogenetic tree based on 16S rRNA sequences (Figure 2.4.1) provided a better measure of rearrangements because its phylogenetic distance was approximately equal to target genomes of *T. scotoductus* SA-01 and *T. thermophilus* HB8 and *T. thermophilus* HB27.

It was theorised that genome rearrangements may disrupt operons by either splitting or relocating genes. Despite observed differences in average operon length, the comparison between the reference and target genomes revealed that the differences in the average operon lengths were not statistically significant. The impact of genome rearrangements on operon length was therefore insignificant. Breakpoints were distributed more or less randomly throughout the whole genome with a bit higher frequency in GC rich regions, but the predicted horizontally transferred genomic islands were less affected by the relocation events (Figure 2.7.3). Genomic islands comprised a small proportion of the entire chromosome which was a limitation on the number of breakpoints that could occur within them. In addition, some genomic islands may have ameliorated to resemble the native genome. Furthermore, more rearrangements could have been triggered by horizontal gene transfer although they could not be directly associated with genomic islands. As such, the overall impact of horizontal gene transfer was inconclusive from the number of breakpoints occurring within genomic islands. However, results show that horizontal gene transfer did not affect the frequency of rearrangements in *Thermus* species which probably could be due to some internal factors associated with the core genome properties. Further hierarchical analysis of horizontal gene transfer with regard to donors and recipients was performed and is presented in Chapter three.

By DNA compositional analysis of genomic fragments of *T. scotoductus* SA-01 flanking the breakpoints, it was found that a 'GCGCGC' motif was almost 4 times more frequent in 40 bp

upstream and downstream of the breakpoints than in the rest of the genome of *T. scotoductus* SA-01 in general (Table 2.7.2 and Figure 2.7.8). These (GC)_n repeats may facilitate homologous recombination between different chromosomal regions. The frequency of the oligonucleotide 'GCGCGC' was counted in non-coding sequences of *T. scotoductus* SA-01 and of *T. thermophilus* HB8 and HB27. They were twice as likely in noncoding regions as compared to the coding region. They were twice as much in *T. thermophilus* strains as compared to *T. scotoductus* SA-01. This observation suggested that in *Thermus* species genome rearrangements frequently occurred between genes and operons making functional disruptions rare. And indirectly, this finding contributes to the hypothesis that the increased frequency of rearrangements in extreme thermophiles of *T. thermophilus* strains was an adaptive mechanism as the natural selection favoured accumulation of poly GC motifs in *T. thermophilus* genomes as compared to *T. scotoductus* SA-01. As these motifs presumably are genome rearrangement recognition sites, their accumulation has increased the rate of rearrangements. Schwarzenlander *et al.* (2009) also reported that thermophilic organisms were characterized by frequent genome rearrangements and an increased genomic plasticity, although no biological explanation of this phenomenon was proposed.

Clustering of common pathways in all organisms (Figure 2.8.1) distinguished thermophiles from mesophiles, with thermophiles being generally poorly clustered. The observation was reinforced by the clustering based on genomic distance for common pathways of all organisms (Figure 2.8.2). In both clusterings, the mesophilic organisms, *E. coli* K-12 and *B. subtilis* strain 168 yielded better clustering as compared to thermophiles. These observations were consistent with earlier findings of poor clustering in genomes with high levels of rearrangements. It remains to be investigated as to whether these rearrangements result into new metabolic capabilities essential for survival in thermal environments.

2.10 Summary

This work investigated whether horizontal gene transfer contributed to genome rearrangements. In addition, it analysed the extent to which genome rearrangements occurred and their implication on the organisation of functionally related genes on the chromosome and the coherence of the metabolic network. Genome rearrangements were analysed firstly, by the distribution of functionally related genes in thermophiles (*T. scotoductus* SA-01, *T. thermophilus* HB8 and *T. thermophilus* HB27); moderate thermophiles (*M. silvanus* DSM 9946 and *M. ruber* DSM 1279 and mesophilic *E. coli* K-12 and *B. subtilis* str. 168). Secondly genome rearrangements were

investigated by computing the total number of breakpoints in *T. scotoductus* SA-01, *T. thermophilus* HB8 and *T. thermophilus* HB27 with reference to *M. silvanus* DSM 9946.

The analysis revealed a uniform gene distribution in all genome categories despite minor differences within the shortest distance category of 0-1000 among individual genomes. There was no significant association between the distribution of functionally related genes and the level of thermostability as defined by the optimum growth temperature. These results showed that genes were distributed in a similar fashion among different bacteria species regardless of their levels of thermostability and the frequency of genome rearrangements. Functionally related genes were located in proximity in thermophilic bacteria just like all other bacteria species suggesting that the gene distribution was constrained by their functional relationship.

Breakpoints were randomly distributed with higher occurrence in *T. thermophilus* HB8 and HB27 as compared to *T. scotoductus* SA-01. A dependency was observed between genome rearrangements and the levels of thermostability as per environmental optimum growth temperature. The variation in the frequency of genome rearrangements also correlated to the phylogenetic relationship among species. Genome rearrangements as measured by breakpoints did not significantly disrupt the operons which make them have less biological meaning on the organism. Horizontal gene transfer affected a minor proportion of the genome which implied their insignificant contribution to rearrangements. Assessment of rearrangements hotspot showed that they were distributed in consistence with the GC content variations throughout the genome with the high density at the beginning of the chromosome. Assessment of breakpoints regions for abundant signature revealed double the occurrence of 'GCGCGC' motifs around breakpoints and in noncoding regions as compared to coding regions.

Poor metabolic network clustering was observed in thermophiles as compared to mesophiles which was consistent with the high levels of genome rearrangements and thermostability in thermophiles. Coherence of metabolic networks was similar to the phylogenetic relationship among organisms suggesting similar evolutionary pressures. Genome rearrangements appear to have considerably shaped the evolution of metabolic networks as evidenced by the distinctions in the clustering levels. Among *Thermus* species, *Thermus scotoductus* SA-01 had fewer rearrangements and yet better clustered than *T. thermophilus* HB8 and *T. thermophilus* HB27 which had more rearrangements. Hence, the environmental optimum growth temperature can be implicated by shaping the structure of genomes and metabolic networks.

Chapter 3

Horizontal Gene Transfer

3.1 Introduction

Horizontal gene transfer (HGT) also known as lateral gene transfer is the acquisition of genetic material from one micro-organism to another. This transfer of genes is different from acquisition of genes through reproduction process which is the vertical transfer of genes between organisms. Apart from mutation and natural selection, horizontal gene transfer is one of the major players in evolution of micro-organisms. Evolution has been predominantly understood through slow processes of mutation and natural selection. However, mutation and natural selection could not explain drastic changes in organisms such as non-pathogenic bacteria becoming pathogenic. Horizontal gene transfer has defied conventional neo-Darwinian theories of evolution to explain evolutionary “quantum leaps” of newer traits in organisms. As opposed to mutation and natural selection which are extremely slow, horizontal gene transfer facilitate huge changes in micro-organisms within comparatively shorter periods of time. Acquired DNA chunks known as genomic islands (GIs) comprise genes that encode novel traits and functions for adaptation but also essential in medicine, agriculture, environment and in industry. Such novel traits and functions include antibiotic and drug resistance, pathogenicity, metal reduction, nutrient fixation among many others.

Horizontal gene transfer is achieved through conjugation, transduction and transformation. These transfer mechanisms depend on the phylogenetic closeness of the micro-organisms. Through conjugation genetic material is transferred via cell to cell contact of conjugal plasmids or transposomes usually between distantly related bacteria. Transduction is facilitated by phages to move genetic material between closely related organisms. Through natural transformations, naked DNA is acquired from the environment from dead cells and incorporated into genomes of micro-organisms. Investigation of horizontal gene transfer among *Thermus* species was motivated by the work of Schwarzenlander *et al.* (2009) and Friedrich *et al.* (2003) where high rates of natural transformation as a major technique for horizontal gene transfer was claimed to be a survival technique in thermal environments. A macro-molecular transport machinery consisting of 16 subunits of pilin-like competence genes that spans the cytoplasmic membrane has been implicated in this process. Further investigations showed that disruption of any of the 16 pilin-like proteins result in considerable negative effect on levels of DNA binding and transportation through the cell membrane (Friedrich *et al.*, 2003).

Several approaches are used to detect horizontal gene transfer. These approaches are based on sequence similarity and nucleotide composition similarity approaches. They detect horizontal gene transfer by searching for homologous sequences. DNA composition methods use genome signatures such as codon or oligonucleotide usage to identify genomic islands. GC composition, codon bias and oligonucleotide features which vary among organisms are used to identify genomic islands. Acquired genomic islands have GC content, oligonucleotide words and codon usage similar to donor and dissimilar from their current host organism. The process of amelioration limits the use of these methods to identify genomic islands because over time genomic islands gradually resemble the composition of the native genome as they get subjected to context dependent mutations, native gene replication, and repair and segregation mechanisms. Genomic islands acquired from closely related species are also difficult to detect because of the similarity of DNA composition between donor and recipient organisms.

Against this background, this work aimed at identifying genomic islands. It investigated their functional contribution towards the adaptation of *Thermus* species in thermal environments. Sources of genomic islands were investigated through the donor-recipient analysis.

3.2 Aim

The aim of this study was to:

- Identify genomic islands and horizontally transferred genes in *Thermus* species.
- Determine the functional contribution of horizontal gene transfer (HGT) in the evolution of *Thermus* species.
- Investigate donor-recipient relationship of genomic islands.
- Estimate the possible time of insertion and the age of genomic islands.

3.3 Materials and Methods

The theory of evolution in which among other processes involves modification of genetic material through mutation, recombination and natural selection has led to the survival, diversification and speciation of organisms reinforced by positive selection which promotes and retains beneficial changes. Phylogenetic relationships between species have been constructed based on these evolutionary molecular markers using the highly conserved 16S rRNA sequences. Conservation of 16S rRNA sequences has proved to be a good marker for constructing phylogenetic relationships among species. However, sometimes construction of phylogenetic relationship among species using other orthologous genes yields different, conflicting and

incongruent relationships. Species which are different morphologically, physiologically and by other molecular markers end up being grouped together and visa versa. This has led to the proposal of potential cross-species gene transfer which is possible through horizontal gene transfer as an alternative explanation of conflicting phylogenetic relationships. Several approaches have been developed to determine distance between genomes due to cross-species gene transfer. Such approaches are based on ortholog frequency and gene order conservation (Wolf *et al.*, 2001; Collyn *et al.*, 2006; Snel *et al.*, 1999). The work of Snel *et al.* (1999) and Collyn *et al.* (2006) drew evolutionary relationship between pathogenicity islands and plasmids by measuring the proportion of orthologous genes between two genomes. The distance between two species was determined by the number of common genes between them divided by the total number of genes. The evolutionary distance in this case is influenced by acquisition and loss of genes between two species. Analysing horizontal gene transfer in this manner explains the movement of genomic islands across species. As these genomic islands move in both directions, what remains is to analyse them in a stratigraphic manner and determine their net movement.

3.3.1 Identification of horizontally acquired genes

Five *Thermus/Meiothermus* genomes were analysed: *T. thermophilus* HB8 and HB27, *T. scotoductus* SA-01, *M. silvanus* DSM 9946 and *M. ruber* DSM 1279. Based on the approach by Collyn *et al.* (2006), an in-house python scripts was implemented. Freely available bioinformatics utilities for the local BLAST and multiple sequence alignment programs were applied. Orthologous genes among the five genomes were identified with the aim of investigating possible gene exchange across species. A BLASTp search was done in a pair wise manner for all coding sequences of five sampled genomes. A total of 1,526 groups of orthologous protein shared by the five genomes were found. Sequences in all groups of orthologous genes were aligned by muscle program. Additionally, the alignments were edited by Gblocks program (Talavera & Castresana, 2007) to exclude badly aligned sequences. For each group of orthologous sequences, phylogenetic trees were constructed using the Neighbour-Joining algorithm implemented in neighbor.exe (Felsenstein, 1993). In every group, a sequence originating from *M. ruber* DSM 1279 was used for rooting. Topological disagreements between gene trees were inspected by the Treedist program (Felsenstein, 1993). Mismatches between trees implied either different rates of mutations in distant taxa, or horizontal gene transfer. It was hypothesized that the differences in the rates most likely affect lengths of branches in phylogenetic trees, while horizontal gene transfer would cause topological changes. An exceptional scenario would occur if a gene in one organism lost its functionality due to nonsense mutations or gene truncation that also may have

resulted in a tree topology alteration. To exclude this situation, only alignments in which unambiguously aligned blocks selected by Gblocks comprised 75% or more of the initial alignment were studied.

Topologies of 1,384 gene trees were identical to the consensus tree. Eleven alternative tree topologies were found to be incongruent to the consensus tree which may be explained by horizontal gene transfer across species or from other unknown sources. The topology in which *M. silvanus* DSM 9946 was clustered together with *T. thermophilus* HB8 and HB27; *T. scotoductus* SA-01 formed a second group with *M. ruber* DSM 1279 with a frequency of 37 trees based on unambiguously aligned sequences. This topology implied a number of possibilities, such as: an exchange of genes between *M. silvanus* DSM 9946 and *T. thermophilus* lineages in any direction; the exchange could be between *T. scotoductus* SA-01 and *M. ruber* DSM 1279 lineages; or by acquisition of diverse genes by *T. scotoductus* or *M. silvanus* DSM 9946 from unknown lineages. To choose the most likely scenario, average relative distances between the 5 genomes was calculated based on 1,384 gene trees sharing the topology with the consensus tree. For normalization, the distances between corresponding nodes in a tree were divided by the total length of all branches of the tree. In trees with alternative topologies, each genome was characterized by the amount of movement of the corresponding node relative to other nodes as given in equation 3.3.1.

$$S_{ij} = \sum |dist_{jk}^i - dist_{jk}^{cons}| \quad (3.3.1)$$

where S_{ij} is the characteristic parameter calculated for species j in tree i ; $dist_{jk}^i$ is the normalized distance between species j and k in given tree i ; and $dist_{jk}^{cons}$ is the distance between the same species in the consensus tree. An organism which gained maximal S_{ij} in the tree was the most likely recipient of horizontally transferred genes, and an organism closer to the recipient organism in the gene tree compared to the consensus tree was a possible donor. If all distances from the recipient organism to others in the gene tree increased, it was assumed that the gene was acquired from an unknown source.

3.3.2 Stratigraphic and donor-recipient analysis

The stratigraphic method calculated distances between oligonucleotide patterns (OU) of genomic islands and host chromosomes to determine relative time of acquisition. Based on oligonucleotide parameters as per the work of Ganesan *et al.* (2008) and Reva *et al.* (2005), a method described by Bezuidt *et al.* (2011) was implemented in this work to analyse stratigraphic and donor-

recipients. Oligonucleotide patterns were used to compare DNA sequences which essentially was a matrix of deviations Δ_w of observed versus expected number of words for all possible words of length N as shown in formula 3.3.2.

$$\Delta_w = \frac{6 \times \left(\frac{\ln(C_{ob}^2) \times \sqrt{C_{exp}^2 + C_0^2}}{C_{exp}^2 \times \sqrt{C_{ob}^2 + C_0^2}} \right)}{\sqrt{\ln(C_0^2 / C_{exp}^2 + 1)}} \quad (3.3.2)$$

where C_{ob} and C_{exp} are observed and expected counts of words w respectively; C_0 assumes equal distribution of words in a sequence. The sum of absolute distances between ranks of identical words after ordering them by their deviation values define the distance D between patterns i and j in formula (3.3.3) and $D(\%)$ denotes the compositional sequence similarity.

$$D(\%) = 100 \times \frac{\sum_w^{4N} |\text{rank}_{w,i} - \text{rank}_{w,j}| - D_{min}}{D_{max} - D_{min}} \quad (3.3.3)$$

Another parameter, pattern skew (PS) defines D calculated for the same sequence where i and j are forward and reverse sequence strands respectively. Based on these values the donor-recipient relationship was computed as per formula (3.3.4).

$$Y = \frac{d_1^2 - d_2^2 + D^2}{2D}; X = \sqrt{D^2 - Y^2} \quad (3.3.4)$$

Where d_1 denotes the distance between host chromosome and the genomic island; d_2 is the distance between the possible donor chromosome and the genomic island while D is the distance between chromosomal oligonucleotide patterns. Distances X and Y were plotted on two dimensional XY plane.

Figure 3.3.1 is an example illustrating graphical representation of compositional relations between genomes and genomic islands of *Shewanella denitrificans* and *Salmonella enterica* strains calculated by the method described by Bezuidt *et al.* (2011). Assume that g_{i1} and g_{i2} are genomic islands identified in *Salmonella enterica* and that *Shewanella denitrificans* respectively. These genomic islands were chosen because their patterns were relatively closer as compared to other genomic islands within the same organisms. This computation is practically possible when genomic islands are over 75% similar or are at least linked by intermediate nodes. To determine the putative donor, the following distances were computed: $D_{I1} = 34\%$, is the distance between the host chromosome and the genomic island g_{i1} ; $D_{I2} = 61\%$ is the distance between the possible donor chromosome and the genomic island g_{i1} . D_{21} is the distance between genomic island and

the possible donor while D_{22} is the distance between the island and the host. $D = 45\%$ is the distance between patterns of the two genomes. From this example, it is hard to conclude that g_{i1} was donated by *Salmonella enterica* because its pattern was relatively closer to its host than to the possible donor. This genomic island may have been donated by another organism whose pattern is similar to that of *Shewanella denitrificans* or indeed the genomic island is very old and has ameliorated over time. It can be concluded however, that *Shewanella denitrificans* is the possible donor of g_{i2} because its pattern is closer to *Shewanella denitrificans* than to its host.

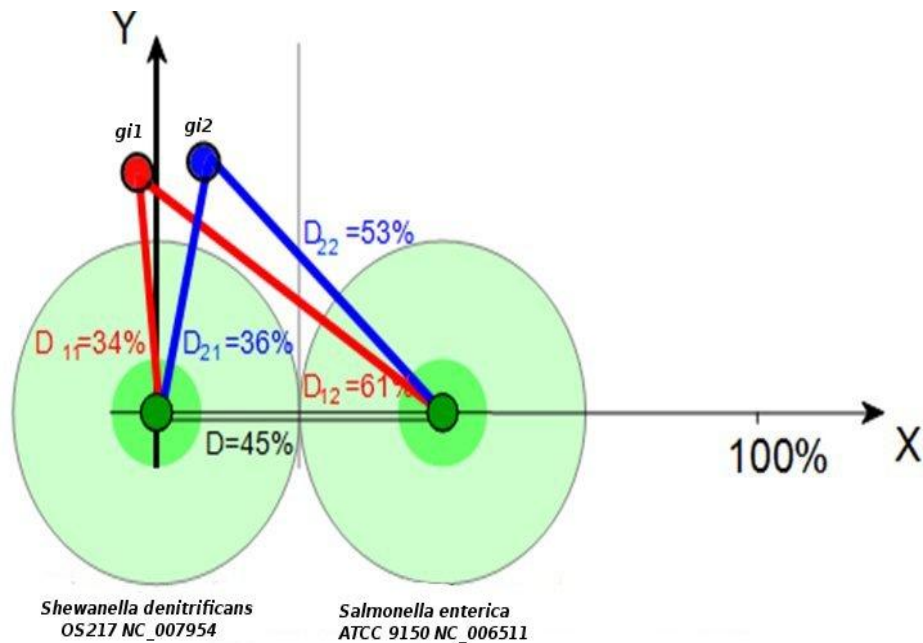


Figure 3.3.1: Stratigraphic donor-recipient illustration (Bezuidt *et al.*, 2011)

The dark green circle in the centre depicts chromosomes which are surrounded by the shaded green area representing intergenomic variability. The intermediate grey shaded line together with the light green areas stand for half distance of chromosomal *OU* patterns. Genomic islands are shown as red and blue circles labelled g_{i1} and g_{i2} . *Y* axis is the *OU* pattern distance between genomic islands and their host chromosomes. The distance between hosts chromosomes are shown by the *X* axis.

3.3.3 Identification and functional analysis of genomic islands

Genomic islands in bacterial genomes were predicted by the SeqWord Genome Browser tool (Ganesan *et al.*, 2008) and its semiautomatic realization SeqWord Snifer (Bezuidt *et al.*, 2009), which are available at the SeqWord project website (www.bi.up.ac.za/SeqWord/). The SeqWord Genome Browser and SeqWord Snifer are tools for identification of genomic islands based on the oligonucleotide (*OU*) patterns. Statistical parameters that characterise the *OU* pattern for

identification of genomic islands discussed in preceding section include: distance between two patterns of the same type (D); pattern skew (PS), is the distance between two patterns of reverse and direct strands; and relative variance (RV). Patterns are calculated and compared for different word length between two genome sequences. Words of size four sliding by one base were used in an 8 kilobases window stepping by 2 kilobases. SeqWord Snifer comprises an underlying database of genomes that were automatically compared against. A critical problem in identification of genomic islands is to determine their precise end points on the genome. SeqWord Snifer was preferred because it is more sensitive in determining borders of genomic islands. Genes and their functions within genomic islands were identified and analysed.

The aim was to identify genomic islands within *Thermus* species and their acquired functionality. Through horizontal gene transfer, bacteria acquire new functionality through genes responsible for virulence, antibiotic resistance and adaptation in general.

3.4 Results and Discussion

3.4.1 Stratigraphic and Donor-Recipient

Horizontally transferred genomic islands were identified in *Thermus* genomes and other species using SeqWord Snifer program (Bezuidt *et al.*, 2011). Genome atlases with indicated positions of genomic islands are available online on (www.bi.up.ac.za/SeqWord). A search through the database of predicted genomic islands in completely sequenced bacterial genomes revealed a compositional similarity of genomic islands found in *Thermus* species to a broad group of mobile genetic elements in *Deinococcus*, *Actinobacteria* and some other bacterial taxa (Figure 3.4.1 A). Mobile genetic elements undergo amelioration which is a process that levels oligonucleotide usage patterns of the acquired genetic elements to resemble their host chromosomes (Lawrence & Ochman, 1997). Figure 3.4.1B shows results of stratigraphic analysis of genomic islands represented as linked nodes. The stratigraphic method calculates distances between oligonucleotide usage patterns of genomic islands and their host chromosomes to determine their relative age.

In Figure 3.4.1B the nodes which are depicted by a lighter colour show higher levels of compositional similarity to the host chromosomes than those depicted by a darker colour, which probably still resemble the composition of their donor genomes. Thus, the colours of nodes are linked with the acquisition periods. Recent acquisitions are distant from the hosts in terms of DNA composition; they therefore have a darker colour. A colour in this case means that these genomic islands have not yet lost their specific original composition. Overtime, these genomic islands get affected by their new hosts through the process of amelioration. Lighter genomic

islands in Figure 3.4.1B are ancient acquisitions, which have been in the host chromosomes for longer, hence the resemblance of the patterns. The stratigraphic analysis demonstrated that prophages of mycobacteria were the most ancient genomic inserts. Later these genomic islands were acquired by *Deinococcus* (most likely *D. geothermalis* lineage) from where they were transmitted to *Thermus*, *Meiothermus* and *Deinococcus* species. But even in *Thermus* species these genomic islands were relatively old as compared to similar inserts in β - and γ -Proteobacteria (Figure 3.4.1B)

3.4.2 Movement of horizontally transferred genes across species

Average relative distances were calculated for 5 *Thermus/Meiothermus* genomes (*T. thermophilus* HB8 and HB27, *T. scotoductus* SA-01, *M. silvanus* DSM 9946 and *M. ruber* DSM 1279). It cannot be excluded in each particular case that an unexpected similarity between orthologous genes of two distant organisms may result from a genetic convergence rather than horizontal gene transfer. However, for the organisms to possess such an elaborated DNA uptake system as *Thermus* species, the hypothesis of the lateral gene exchanges was plausible. Even if not all these genes were horizontally acquired, this analysis demonstrated which organisms tend to share more of their genetic material and which of them are inclined to be recipients of DNA fragments. A summary of the most common donor-recipient links is shown in Figure 3.4.2. The sizes of the arrows are proportional to the movement and the number of genes from one organism to another.

All mobile genes were identified in chromosomal location in both donors (if known) and recipients. It was found that *T. scotoductus* SA-01 frequently acquired DNA fragments from *Meiothermus* species than vice versa. It is quite possible that the relatively thermophilic *Meiothermus* species and *T. scotoductus* SA-01 share their habitats than with the extremely thermophilic *T. thermophilus* strains. *T. scotoductus* SA-01 also acquired foreign DNA, mostly from unknown sources. The capability of *Meiothermus* species to uptake DNA fragments is noticeably weaker than that of *Thermus* species. Genes were possibly acquired by DNA transformation encoded for ribosomal proteins, enzymes of amino acid biosynthesis and other metabolic pathways. In contrast to genomic islands, which may bear genetic clusters encoding whole pathways, short DNA fragments comprise only one or few genes. Thus, the conserved genes involved in the basic metabolic functions stand a better chance to be used in a new host and to persist over generations. Replacement of own genes by alternative foreign variants may be advantageous for fine-tuning and timing of biological processes and protein-protein interactions to adapt to changing environmental conditions.

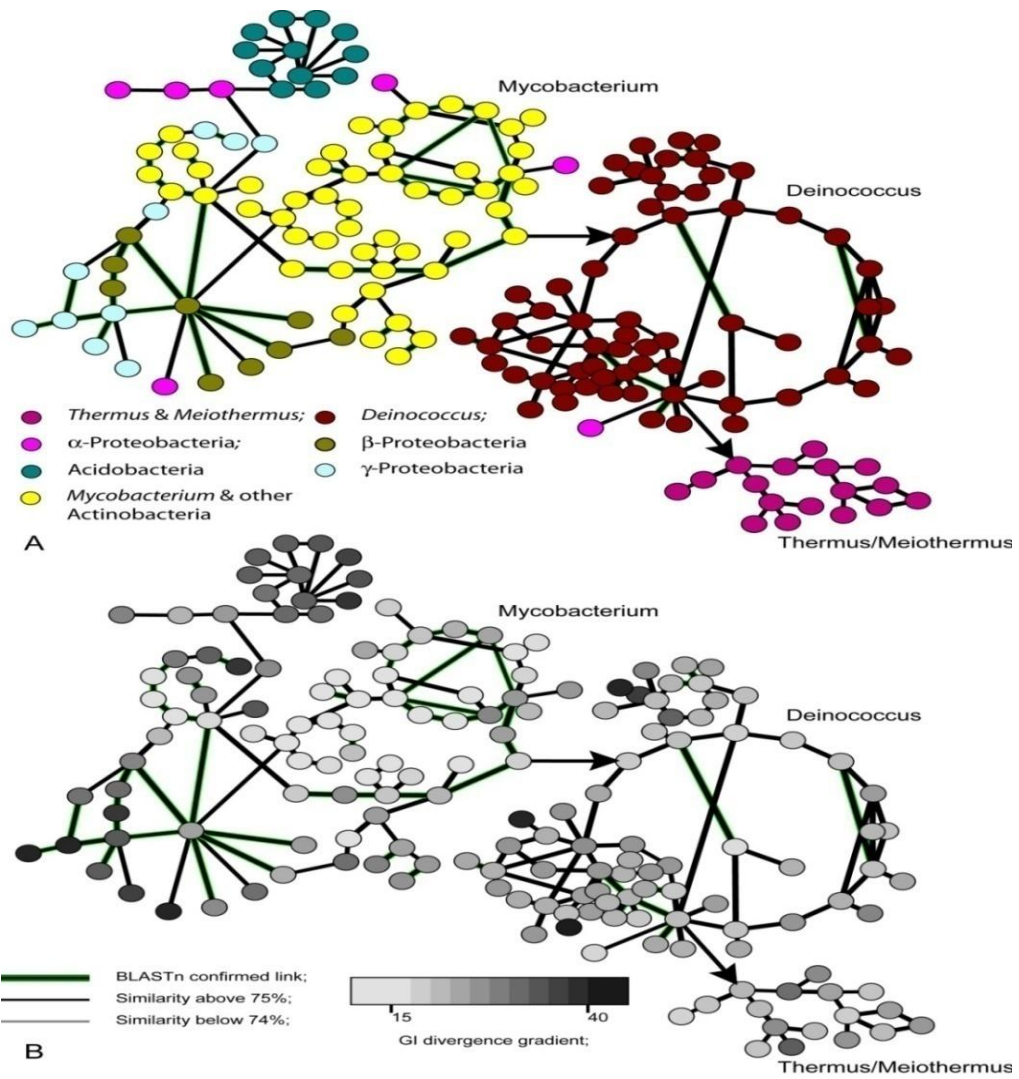


Figure 3.4.1: Age of genomic islands and relative time of insertion

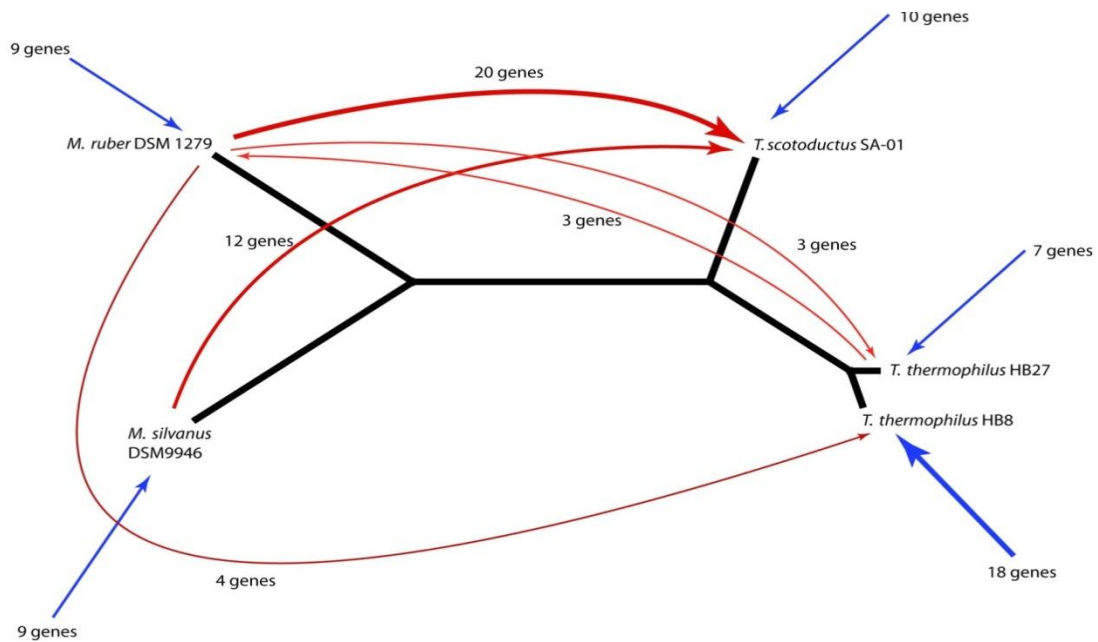


Figure 3.4.2: Acquisition and movement of genomic islands across species

3.4.3 Functional analysis of genomic islands

SeqWord Snifer program allowed predicting of relatively large inserts of foreign DNA, mostly hypothetical genes, prophages and fragments of conjugative plasmids. Table 3.4.1 provides details of total number of predicted genomic islands; number of genes and their functions; including total number of pathways involved.

Table 3.4.1: Analysis of genomic island (GIs) as predicted by SeqWord Snifer for *T. scotoductus* SA-01, *T. thermophilus* HB8 and HB27

Genomic Islands(GI)				Type of Pathways and Genes within GIs		
Genome	GIs	Genes	Hypothetical	Others	Degradation	Biosynthesis
SA-01	6	169	83	5	8	5
HB8	2	47	31	0	0	1
HB27	4	110	37	6	1	12

The table shows number of genomic islands identified in *Thermus* organisms using various genomic island detection programs. It also shows the number of both genes and hypothetical genes within genomic islands of each organism. Then, for all genes it checks the pathways in which they act and presents the number of pathways related to horizontal gene transfer for each organism.

In addition to other acquired novel functions in genomic islands, most genes coded for enzymes catalysing biosynthesis and degradation pathways. A large number of horizontally acquired enzymes in *T. scotoductus* SA-01 catered degradation pathways (8) such as 4-hydroxymandelate, biphenyl, protocatechuate, ethanol, 2-oxopentenoate, 2-aminophenol and oxidative ethanol. Biosynthesis pathways (5) found in genomic islands included: O-antigen, hexaprenyl diphosphate, isoleucine, menaquinone-8 and adenosylcobalamin from cobyrinate A,C diamide II. Prior work has revealed that the K and O-antigens play a big role in resistance to host defences (Roycroft *et al.*, 2006). Majority of genes within genomic islands were involved in cell wall polysaccharides biosynthesis which contribute to bio-film formation and cell surface O-antigen variability, consistent with previously reports that this category of enzymes were abundant in mobile genetic elements (Bezuidt *et al.*, 2011).

Several genes encoding membrane spanning protein responsible for moving ions, molecules and macromolecules were horizontally acquired. These aid movement of DNA molecules from environment through cell membranes to facilitate natural transformation. Cell walls play a major role in thermal and toxic inorganic compound resistance in thermophilic bacteria such as resistance to potassium tellurite in *T. thermophilus* HB8 and *T. favus* AT62 (Chiong *et al.*, 1988).

Genes coding for transposase and the IS4 family protein were also horizontally acquired in *T. scotoductus* SA-01. The IS4 protein family are small mobile genetic entities that enable expansion of prokaryotic genomes and trigger important rearrangements (Palmenaer *et al.*, 2008; Aziz *et al.*, 2010). In *T. thermophilus* HB8 one enzyme, putative phosphoglu-

comutase/phosphomannomutase in the GDP-glucose biosynthesis pathway was identified in genomic islands. In *T. thermophilus* HB27, however, most enzymes coded for biosynthesis pathways (12) such as the super pathway of leucine, valine and isoleucine, asparagine, CMP-N-acetylneuraminic acid and GDP-glucose biosynthesis. Genes for oxidative ethanol degradation (also in genomic islands of *T. scotoductus* SA-01) and a nitrate reduction pathway were found in *T. thermophilus* HB27 genomic islands both essential for biotechnology application. *T. scotoductus* SA-01 is shown to use NO_3^- as a terminal electron acceptor for growth in the absence of oxygen (Kieft *et al.*, 1999). *Thermus* species can therefore contend in bio-fuel production such as bio-ethanol with their ability to function in high temperature under anaerobic conditions. Bio-fuels are produced in enclosed environment where temperatures rise due to decomposition as oxygen levels deplete.

Genes coding for pilin-like proteins, *pilA*, *comZ* and competence protein *dprA* implicated in aiding natural transformation were horizontally acquired in *T. thermophilus* HB27. Both *pilA* and *dprA* are involved in DNA binding and transportation through the outer membrane. *ComZ* is more involved in DNA transportation during natural transformation (Schwarzenlander *et al.*, 2009). In *T. scotoductus* SA-01 and *T. thermophilus* HB8, pilin-like protein coding genes occur in the chromosome although they were not detected to be horizontally transferred. It may be that they have ameliorated over time to resemble the native genome composition. Several genes coding transposases aid movement of transposons to other parts of the genome (Gómez *et al.*, 2012). These were identified within genomic islands of both *T. thermophilus* HB8 and *T. thermophilus* HB27. A large number of horizontally transferred genes were annotated as conserved hypotheticals, impeding inference of possible roles that these genomic islands might have played in the evolution of *Thermus* species. Majority of genes in all the three genomes were hypotheticals with unknown function. Complete annotation to elucidate their functions would shed more light on horizontally acquired novel function and their role in thermal adaptation.

3.5 Summary

Stratigraphic analysis of genomic islands (Figure 3.4.1) showed movement of genomic islands from *Mycobacterium* to *Deinococcus* and then to *Thermus/Meiothermus*. This confirmed the ability of *Thermus/Meiothermus* species to acquire new genetic material suggesting possible novel acquired functions related to thermal adaptation that could have potential biotechnology application. Relating this observation to the donor-recipient analysis, within the *Thermus/Meiothermus* species, there was more movement of genes in relative terms from *Meiothermus* to *Thermus* species, further suggesting an established ability of DNA uptake

mechanism in *Thermus* species as compared to *Meiothermus*. More genes were acquired by *T. scotoductus* SA-01 than *T. thermophilus* strains from *Meiothermus* species because of the possibility of higher interaction between *T. scotoductus* and *Meiothermus* species since they inhabit environments with relatively closer levels of temperatures. These results suggested that DNA exchange was not limited to natural transformations in *Thermus* species, but other mechanisms also contributed such as conjugation and transduction. *T. thermophilus* according to this analysis acquired more genes from other sources than fellow *Thermus/Meiothermus* species. It is not surprising to observe fewer rearrangements in *T. scotoductus* SA-01 as compared to *T. thermophilus* since *T. scotoductus* SA-01 acquired majority of those genes from *Meiothermus* species from which the reference genome was selected. Gene synteny was not adversely affected by horizontal gene transfer since their order may have followed that of the ancestral genome. Genes acquired elsewhere by *T. thermophilus* could more significantly contribute to rearrangements than those acquired from *Meiothermus* species.

The contribution of genomic islands to genome rearrangements was found to be insignificant. The number of genomic island regions was proportionately much smaller in comparison to the entire genome. However, functional analysis of acquired genes within genomic islands suggested that horizontal gene transfer has much to do with genome rearrangements. In *T. thermophilus* HB27 for example, genes coding for pilin-like proteins implicated in natural transformation were found to be horizontally acquired in *T. thermophilus* HB27. Several transposases and intergrases responsible for genome structure modification by moving DNA sequences within and between genomes were horizontally acquired in *T. scotoductus* SA-01 and *T. thermophilus* HB8. Although these genes legitimise detected genomic islands, their implication on chromosomal rearrangements is beyond the genomic islands.

Chapter 4

Protein and Enzyme Thermostability

4.1 Introduction

Thermostability is the resistance of a substance to irreversibility of chemical or physical structural changes due to increase in temperature. Protein or enzyme thermostability is therefore, the preservation of the unique structure and chemical properties of polypeptide chains under high temperatures (Zhou *et al.*, 2008). Thermostable enzymes enhance productivity in industry as they operate at higher temperatures where most reagents and compounds are available. Reactions with thermostable enzymes at higher temperatures are more efficient because of more solubility which does not require frequent cooling as is the case with mesophilic or synthetic enzymes. Mesophilic enzymes denature at higher temperatures. Hence, they need frequent cooling using expensive machines to keep their temperatures lower. However, at lower temperatures reactions are slower due to higher viscosity. Solubility is lower and levels of contamination are more abundant in cooler temperatures which increase production costs. Industrial application of thermostable enzymes results in low production costs due to reduction in wastage partly caused by contamination of mesophilic micro-organisms which are abundant in lower temperatures. The elimination of expensive cooling systems reduces production costs and consequently results in higher profits (Haki & Rakshit, 2003). Although some thermostable enzymes are synthesised or genetically engineered through mutagenesis, thermophilic micro-organisms provide a better natural resource. Enzymes are extracted from cultured thermophilic organisms or their properties are elucidated for synthetic production. Micro-organisms can also be directly inoculated on mediums to achieve desired by-products through biochemical processes such as respiration, oxidation and degradation (Seeliger & Groot, 2010). Examples of such application of micro-organisms include: bioremediation through which heavy metal pollution is eradicated, cleansing contaminated water sources, clearing clogged pipes, bio-mining, production of bio-fuels, controlling methane fluxes into the atmosphere which reduces global warming, among many others. *Thermus* species are potential sources of thermostable enzymes that can be applied in industry.

The ability for *Thermus* species to live in higher temperature environments such as hot springs, domestic hot water, deep mines and compost manure (Munster *et al.*, 1986; Williams *et al.*, 1995) is extremely important in biotechnological application (Lasa & Berenguer, 1993). Sustaining life at higher temperatures suggests an evolved metabolic network system and the use

of thermostable enzymes that function without denaturing to facilitate cellular biochemical processes necessary for survival. Such enzymes are both thermostable and resistant to chemical reagents, salt concentrations, high pressure, extremely acidic and alkaline conditions making them even more suitable for biotechnology application (Niehaus *et al.*, 1999). Several factors have been implicated in stabilising proteins and enzymes in thermophilic organisms such as: amino acid preference, ratio of charged versus uncharged amino acids, amino acids preference, ionic interactions, codon usage, hydrophobicity, protein surface area and many others (Kumar *et al.*, 2000; Russell *et al.*, 1997; Zhou *et al.*, 2008). Thermal stability however, is a consequence of the combination of several factors in an additive manner but not due to a single specific major factor (Lasa & Berenguer, 1993; Vogt *et al.*, 1997). Moreover, stabilising factors differ from taxon to taxon, organism to organism and even within the same organism from protein to protein which makes the general elucidation of thermostable enhancing factors extremely complex (Trivedi *et al.*, 2006). It is not surprising that studies have come up with conflicting results on factors that generally influence protein thermostability. Conflicting results have been aggravated by insufficient data to be generalised to all proteins and different methods applied in identifying thermostability enhancing factors.

Among several other approaches, protein thermostability is measured using optimum growth temperature (OGT) (Huang *et al.*, 2004) protein melting temperature (T_m) (Kumar *et al.*, 2000) minimum folding energy (MFE) of RNA secondary structures (Zuker & Stiegler, 1981; Markham & Zuker, 2008), codon preference and the (E+K)/(Q+H) ratio (Fariás & Bonato, 2003). The ratio was determined based on optimum environmental growth temperature of micro-organisms. The optimum growth temperature is a crude estimate that does not reflect practical living conditions of organisms. The same organisms can live in environments with different levels of temperature depending on salinity, acidity and several other environmental factors. The terms, 'environmental temperature' and 'optimum growth temperature' have been used loosely throughout this discussion. Optimum growth temperature may not precisely measure thermostability of individual proteins to be generalised to entire species. Moreover, thermostability differs from protein to protein even within the same organism with the same environmental temperature. Environmental growth temperature cannot be relied on to distinguish thermostability of proteins even within the same organism and let alone to identify individual thermostable proteins. The conventional approach has been to crudely consider all proteins from thermophilic organisms as thermostable. In this work however, the aim was to determine factors that can be used to uniquely identify specific thermostable proteins within *Thermus* species.

Melting temperature is another approach that is used to deduce protein thermostability.

Proteins are physically heated to a level where they denature. The temperature at which proteins denature defines their thermostability. However, physically heating proteins requires advanced expertise, equipment and other resources which make the entire process extremely laborious and expensive. In addition, heating proteins in the entire genome is not only laborious, time consuming and costly, but is also impractical. It is important therefore, to explore *in silico* approaches which can reduce this burden and narrow down thermostable protein candidates that could then be experimentally verified. Minimum folding energy (MFE) despite its limitations is a preferred *in silico* approach for this purpose and has been applied in this work.

4.1.1 Aims

The aim of this study was to determine factors that influence thermostability in *Thermus* species. This aim was achieved through the following specific objectives:

1. To examine *in silico* approaches for measuring protein thermostability in order to identify a suitable approach applicable in *Thermus* species.
2. To analyse amino acid property composition and substitution that enhance thermostability in *Thermus* species.
3. Assess protein structural differences as affected by amino acid substitutions that enhance protein thermostability.

4.2 Materials and Methods

4.2.1 Dataset

Seven genomes categorised according to their levels of thermostability as defined by their environmental optimum growth temperature were analysed:

- Hyperthermophiles: *Pyrococcus abyssi* (NC_000868), *Pyrococcus horikoshii* OT3 (NC_000961)
- Thermophilic: *Thermus scotoductus* SA-01 (NC_014974); *Thermus thermophilus* strain HB8 (NC_006461) and *Thermus thermophilus* strain HB27 (NC_005835); *Thermoplasma volcanium* GSS1 (NC_002689)
- Moderate thermophilic: *Meiothermus silvanus* DSM 9946 (CP002042) and *Meiothermus ruber* DSM 1279 (NC_013946)
- Mesophilic: *E. coli* K12 substr. MG1655 (NC_000913) and *Bacillus subtilis* str. 168 (NC_000964)

4.2.2 Determination of protein thermostability

The objective was to identify an *in silico* approach for measuring thermostability based on which preferred amino acid composition and substitution could be analysed. The (E+K)/(Q+H) ratio, codon usage and minimum folding energy were investigated under this objective. Since these methods were applied on data sets of other organisms, it was necessary to investigate their applicability in *Thermus* species. Conflicting results on factors that enhance protein thermostability in different studies are partly due to limited data and different methods that are used. Hence, these methods and factors could not be generalised to *Thermus* species. Some approaches are suitable for determining factors in closely related organisms while others yield better results in distantly related organisms. It was therefore, necessary to investigate and determine the appropriate method and consequent factors for measuring thermostability in *Thermus* species.

4.2.2.1 Amino acid preference and codon usage

The (E+K)/(Q+H) ratio and codon usage as proposed by Fariás and Bonato (2003) were investigated if they could effectively measure thermostability in *Thermus* species. High preference in hyperthermophiles and thermophiles of AGR encoding for arginine as opposed to CGN codons usage in mesophiles were observed in prior studies (Fariás & Bonato, 2003). This was proposed as a signature for classifying organisms into different levels of thermostability. The (E+K)/(Q+H) was derived from an increased occurrence of glutamic acid (E) and lysine (K), and a decrease in the proportion of glutamine (Q) and histidine (H) in correlation to optimum growth temperature. The ratio was determined to be between 3.2 and 4.5 in thermophiles; less than 3.2 in mesophiles and greater than 4.5 in hyperthermophiles. The data set analysed comprised hyperthermophilic, thermophilic, moderate thermophilic and mesophilic genomes based on their optimum growth temperatures as shown in Table 4.2.1.

Table 4.2.1: Genomes analysed for codon usage and the (E+K)/(Q+H) ratio classified according to their optimum growth temperature

Category	Organism	OGT(°C)
Hyperthermophiles	<i>Pyrococcus abyssi</i>	96 – 98
	<i>Pyrococcus horikoshii</i>	
	<i>Thermoplasma volcanium</i>	60
Thermophiles	<i>Thermus scotoductus</i>	60
	<i>Thermus thermophilus</i>	60-80
Mesophiles	<i>Escherichia coli</i>	35-37
	<i>Bacillus subtilis</i>	

The $(E+K)/(Q+H)$ ratio and codon usage were calculated for all coding sequences in all genomes. The $(E+K)/(Q+H)$ ratio values were sorted from largest to smallest. The expectation was that the ratio would classify genomes according to their optimum growth temperatures. A similar analysis was conducted in all genes to verify codon usage as a classification for organisms into different levels of thermostability. High usage of CGR as opposed to AGR was expected in thermostable genes within the same organism as determined by the $(E+K)/(Q+H)$ ratio.

4.2.2.2 Minimum Folding Energy

RNA structures that fold with the least amount of energy are believed to be more stable as they are less affected by external forces (Mohsen *et al.*, 2010). RNA secondary structures are the major determinants of the final translated protein tertiary structures as forces that determine RNA folding are much stronger than those that apply on tertiary structures. Minimum folding energy was computed using an extended implementation of UNAFold algorithm (Markham & Zuker, 2008), an efficient version of mfold algorithm (Zuker & Stiegler, 1981). The original version takes a single DNA or RNA sequence at a time as input; then it computes energy values and stores them in a text file. Another script is run to extract energy values from the output file of the first script. This is prohibitively slow to compute energy values for all coding sequences in the entire genome, and even much more for several genomes for comparative analysis purposes. The extended UNAFold implementation simplifies this task by taking all coding sequences in fasta or genbank format as input to calculate values in one run. If the input is originally in genbank format, it is parsed to fasta format before running it in the script. The pipeline takes care of inefficiencies encountered when running the original UNAFold program. This saves time that is wasted due to pausing of the scripts to issue instructions on the command line. Folding energy values were negated to convert them to positive values for normalisation by sequence length to eliminate sequence length dependency. Minimum folding energy values were computed for thermophilic, moderately thermophilic and mesophilic genomes listed as per below.

- Thermophiles: *Thermus scotoductus* SA-01 (NC_014974); *Thermus thermophilus* strain HB8 (NC_006461) and *Thermus thermophilus* strain HB27 (NC_005835)
- Moderate thermophiles: *Meiothermus silvanus* DSM 9946 (CP002042) and *Meiothermus ruber* DSM 1279 (NC_013946)
- Mesophiles: *E. coli* K-12 substr. MG1655 (NC_000913) and *Bacillus subtilis* str. 168 (NC_000964)

4.2.3 Composition and substitution of amino acid properties

Protein sequences were analysed to identify preferred amino acid residues and properties that enhance thermostability in *Thermus* species. Orthologous sequences for *T. scotoductus* SA-01 were identified in *T. thermophilus* HB27 using BLAST program. Only orthologous sequences which had over 80% coverage with e-values lower than 0.0001 were selected. In addition, selected orthologous sequences had a difference of less than -30 kcal in minimum folding energy on normalised values. Amino acid composition and substitution analysis was performed on closely related species of *T. thermophilus* HB27 and *T. scotoductus* SA-01. This was to eliminate the influence of other evolutionary factors that could exist in distantly related organisms. *T. thermophilus* HB27 was chosen because it shares more homologous sequences with *T. scotoductus* SA-01 with higher sequence identity than *T. thermophilus* HB8. Analysis of proteins from closely related species allowed for a possible conclusion that differences in the amino acid substitution and composition were more likely due to differences in levels of thermostability and not due to other evolutionary factors.

Occurrence of amino acid residues was calculated for 500 orthologous sequences with highest minimum folding energy difference between *T. scotoductus* SA-01 and *T. thermophilus* HB27. First, normality of the amino acid residue distribution was tested using the *Shapiro-Wilk* normality test. Then, based on whether the residues were normally distributed or not, a parametric *t-test* or nonparametric *Wilcoxon t-test* was applied to determine statistically significant difference in distribution of amino acid residues.

Next, was to analyse the distribution of amino acid properties. Frequencies of amino acid properties were derived from residues from the same sample of sequences. These were 500 sequences with the largest difference in minimum folding energy. Percentages of amino acid properties were calculated and normalised by their sequence length. The average amino acid property distribution for each genome was calculated. Normality for each property distribution was tested by the *Shapiro-Wilk* normality test. The parametric *t-test* for normal distribution or the nonparametric *Wilcoxon t-test* for properties that were not normally distributed was performed for each amino acid property to examine whether differences in distribution between highly and less thermostable sequences were statistically significant at 95% level of confidence.

More importantly, was to determine properties that were gained or lost as amino acids substituted from high to less thermophilic protein sequences. To investigate the role of amino acid substitutions in adaptation of *T. thermophilus* HB27 proteins to high temperature environment, proteins with the largest minimum folding energy differences were analysed. An amino acid substitution table was computed for all residues between selected *T. scotoductus* SA-

01 proteins and their orthologous sequences in *T. thermophilus* HB27. The expectation was that the absolute values of amino acid substitutions were proportional to the frequencies of the amino acids in the sampled proteins. The frequency values for each amino acid were computed and normalised as per formula 4.2.1.

$$X_{ij}^{norm} = \frac{N \times X_{ij}}{f_i f_j} \quad (4.2.1)$$

where X_{ij} is the computed value in the table; f_i and f_j are the corresponding frequencies of amino acids; and N is the total number of residues in the whole alignment of 500 selected proteins. The difference between the direct and reverse substitutions for pairs of amino acids calculated on the distribution of normalized values that deviated beyond the 1.96σ threshold, were considered statistically reliable. Amino acid properties that changed between higher and less thermophilic sequences were deduced from the table.

Correlation coefficients were calculated between the amount of MFE changes and frequencies of substitutions in nucleotide and amino acid sequences of orthologous proteins. Statistical significance of the calculated correlation coefficients was confirmed by checking that the standard t -student parameter (1.96 for $p = 0.05$) was significantly smaller than the estimated ones calculated by the following equation 4.2.2:

$$t_{st} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (4.2.2)$$

where r is a correlation coefficient and n is the number of orthologous genes.

4.2.4 Protein structure homology modelling

Despite conflicting results, several factors at protein structure level have been associated to thermostability. These factors include protein surface area, proportion of cavities, deletion and shortening of loops, buried versus exposed surface areas and many others. As these factors are inconclusive in most studies and they differ from organism to organism, it was necessary to investigate them in *Thermus* species. Protein structures were modelled to analyse differences due to amino acid substitutions which may have effect on thermostability. The process of constructing 3D protein models comprised four steps: (i) target identification and selection of templates, (ii) multiple sequence alignment of the target sequence with its homologs, (iii) construction of models using the template and the alignment and (iv) evaluation and refining the model (Martin-Renon *et al.*, 2000).

Analysis was performed on orthologous sequences of *T. thermophilus* HB27 and *T. scotoductus* SA-01 with differences in sequence length not greater than 50 nucleotides; an alignment quality of greater than 30%; sequence identity of greater than 50% and with largest minimum folding energy difference on normalised values. Protein sequence pairs that met this selection criterion were used as target sequences to search for 3D protein structure templates based on which model structures were constructed for further analysis.

Using each target protein sequence, 3D structure templates were searched using *HHpred* program (Soding *et al.*, 2005) in protein data bank (PDB) (Berman *et al.*, 2000) that were used for modelling. *HHpred* is an online interactive program for protein homology detection and structure prediction using Hidden Markov Models (HMM). The program performs searches using sequences or multiple sequence alignments as input in a wide range of databases such as PDB, SCOP, Pfam, SMART, COG, CDD and many others. In this case however, the PDB was searched for protein structures with all other parameters set to default. A template structure with best scores for both target sequences were used for modelling structures for the two target sequences. Template structures with sequence identity of greater than 30% and resolution of less than 3Å were selected. The template structure with highest quality was validated by Procheck, Anolea, and Qmeans6 programs within the Swiss-Model suite (Arnold *et al.*, 2006). The PDB text file was edited so that the starting position of the atoms corresponded to the structure sequence; the HETATM water atoms were also truncated accordingly. PyMOL (DeLano, 2002) was used to inspect and ensure that there were no loop breakages in the structure. Only chain A was used for modelling. Homologous sequences were searched in the NCBI BLAST public database using each target sequences. Homologous sequences for each target with query coverage greater than 80% and e-values lower than 0.0001 qualified for alignment. Homologous sequences together with the 3D template structure were subjected to structural alignment using different programs such as MAFFT (Kato *et al.*, 2002), Muscle (Edgar, 2004) and Promals3D (Pie *et al.*, 2008) in order to establish the best alignment for modelling. The best alignment was selected, examined and further refined before modelling protein structures using MODELLER v9.11 program (Sali & Blundell, 1993). For each template, there were 100 model structures that were generated from which the best three models were selected based on DOPE Z scores (Shen & Sali, 2006). Quality assessment was further made on the best three models using Procheck, Anolea and Qmeans6 found within the Swiss-Model suite, an integrated workspace of protein structure quality assessment programs. The constructed best model was refined by adjusting the alignment and refining loops. Figure 4.2.1 illustrates the modelling process.

Protein structures were modelled using orthologous sequences which met the selection criteria in which the effect of dominant amino acid substitutions was investigated. A pair of orthologous sequences with differences in thermostability was aligned to identify amino acid substitutions between them. *T. thermophilus* HB27 target sequences yielded high sequence identity matches with templates in PDB as compared to *T. scotoductus* SA-01.

This was possibly because *T. thermophilus* HB27 was sequenced much earlier and has therefore been widely studied. Instead of using the modelled structures, the original structures that matched with *T. thermophilus* HB27 in protein data bank were analysed. Stabilising mutations were introduced into structures and the energy changes were observed.

FoldX plugin for Yasara program version 1.4.21 (Schymkowitz *et al.*, 2005; Durme *et al.*, 2011) was used to calculate the energy difference $\Delta\Delta G(\text{change})$ between the original structure (wild type - WT) and the structure in which a mutation was introduced (mutant - MT). FoldX program predicts energy change between two structures based on introduced mutations. The energy difference is given by $\Delta\Delta G(\text{change}) = \Delta G(\text{MT}) - \Delta G(\text{WT})$. The actual computed energy values for the wild type $\Delta G(\text{WT})$ and the mutant $\Delta G(\text{MT})$ are meaningless; hence they cannot be used to compute structure thermostability. However, it is the energy change $\Delta\Delta G(\text{change})$ values which have been experimentally validated. Stabilising mutations which bring about energy changes are less than zero $\Delta\Delta G(\text{change}) < 0$ while destabilising mutations have energy change values greater than zero $\Delta\Delta G(\text{change}) > 0$. FoldX estimation has an error margin of $\pm 0.5\text{kcal/mol}$, implying that energy values above the error margin have significant stabilising effect.

4.3 Results and Discussion

4.3.1 Codon usage and the Fariás-Bonato ratio

The suitability of Fariás-Bonato ratio to measure thermostability in *Thermus* species was assessed. It was observed in previous studies that hyperthermophiles and thermophiles prefer AGR codons when coding for arginine as opposed to CGN in mesophiles. This signature was analysed and contrasted against Fariás-Bonato ratio. Figure 4.3.1 shows relationship between codon usage and the $(E+K)/(Q+H)$ ratio in seven genomes. Indeed, higher AGR and lower CGN occurrence in hyperthermophiles was observed. However, this distinction was not clear between mesophiles and thermophiles. *B. subtilis* str 168 with optimum growth temperature of $35\text{-}37^{\circ}\text{C}$ has values between *T. volcanium* GSS1 and *T. scotoductus* SA-01 which have about 60°C optimum growth temperature. The ratios of *T. thermophilus* HB8 and HB27 are between *E. coli*

K-12 and *T. scotoductus* SA-01. The (E+K)/(Q+H) ratio however, showed a general correlation to the optimum growth temperature across species but was not consistent in classifying the organisms. For example, *T. scotoductus* SA-01 was classified among mesophiles because it had a lower thermostability with an average ratio of 2.8 which was much lower than the prescribed threshold for thermophiles. *T. thermophilus* HB8 and *T. thermophilus* HB27 had average ratios of 3.41 and 3.35 respectively. These were correctly classified among thermophiles. *T. thermophilus* depicted higher levels of thermostability than *T. scotoductus* SA-01 consistent with their environmental living temperature.

The (E+K)/(Q+H) ratio was analysed in all genes to identify thermostable genes. The expectation was that highly thermostable genes with higher (E+K)/(Q+H) values would give higher AGR and lower CGN when coding for arginine. It was observed that mesophiles (*E. coli* K-12) and thermophiles (*T. scotoductus* SA01 and *T. thermophilus* HB27) both prefer CGR as opposed to AGR codons when coding for arginine (Figure 4.3.2).

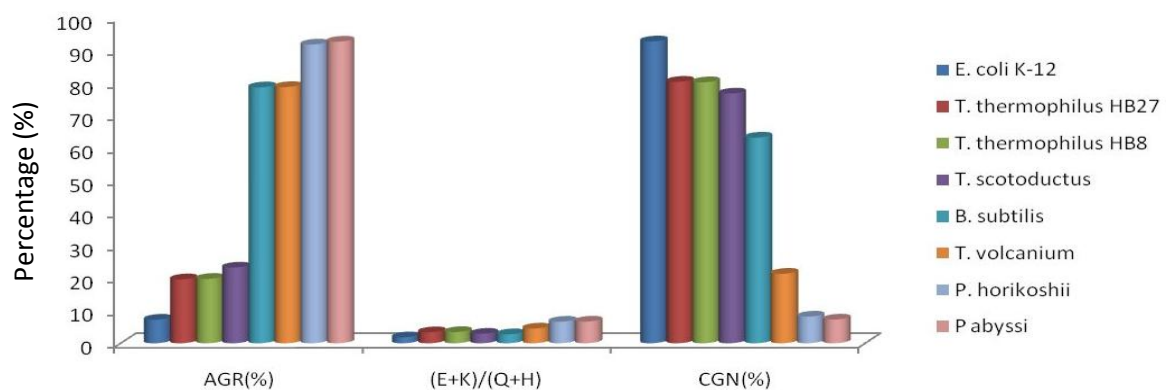


Figure 4.3.1: Relationship between codon usage and the (E+K)/(Q+H) ratio in genomes

Hence, *Thermus* species could not be distinguished from mesophilic based on usage of these codons. However, a clear codon usage distinction was visible in hyperthermophile *P. horikoshii*.

It was also deduced from the computation that as the (E+K)/(Q+H) ratios increased, differences in the AGR and CGN codon usage reduced considerably. This could be used to identify individual thermostable proteins sequences. The (E+K)/(Q+H) ratio was calculated for all genes in *T. scotoductus* SA-01, *T. thermophilus* HB8 and *T. thermophilus* HB27 to identify genes with high levels of thermostability (Figure 4.3.3). The computation excluded heat shock proteins such as chaperons and tRNAs which have unrealistically higher levels of the (E+K)/(Q+H) ratio. Values were sorted from the largest to the smallest ratio. The challenge was that all genes had the same level of thermostability derived from the organism's environmental

optimum growth temperature. As such, verification of thermostability at individual gene level based on $(E+K)/(Q+H)$ ratio did not amount to any useful interpretation.

The AGR and CGR codon usage signatures classify thermophiles together with mesophiles. Hence, it cannot be used to identify individual thermostable genes in genomes. Although the $(E+K)/(Q+H)$ ratio provides an average thermostability measure consistent with environmental optimum growth temperature according to Figure 4.3.3, its classification was not consistent with prescribed thresholds. The $(E+K)/(Q+H)$ ratio was therefore also not ideal in estimating thermostability of individual genes for *Thermus* species. While this ratio may be applicable in distantly related organisms, it is not ideal for closely related organisms. As such, other approaches such as the minimum folding energy was explored as a possible measure of thermostability based on which other *in silico* factors could be determined. Minimum folding energy for RNA folding was examined for measuring thermostability *in Thermus* species.

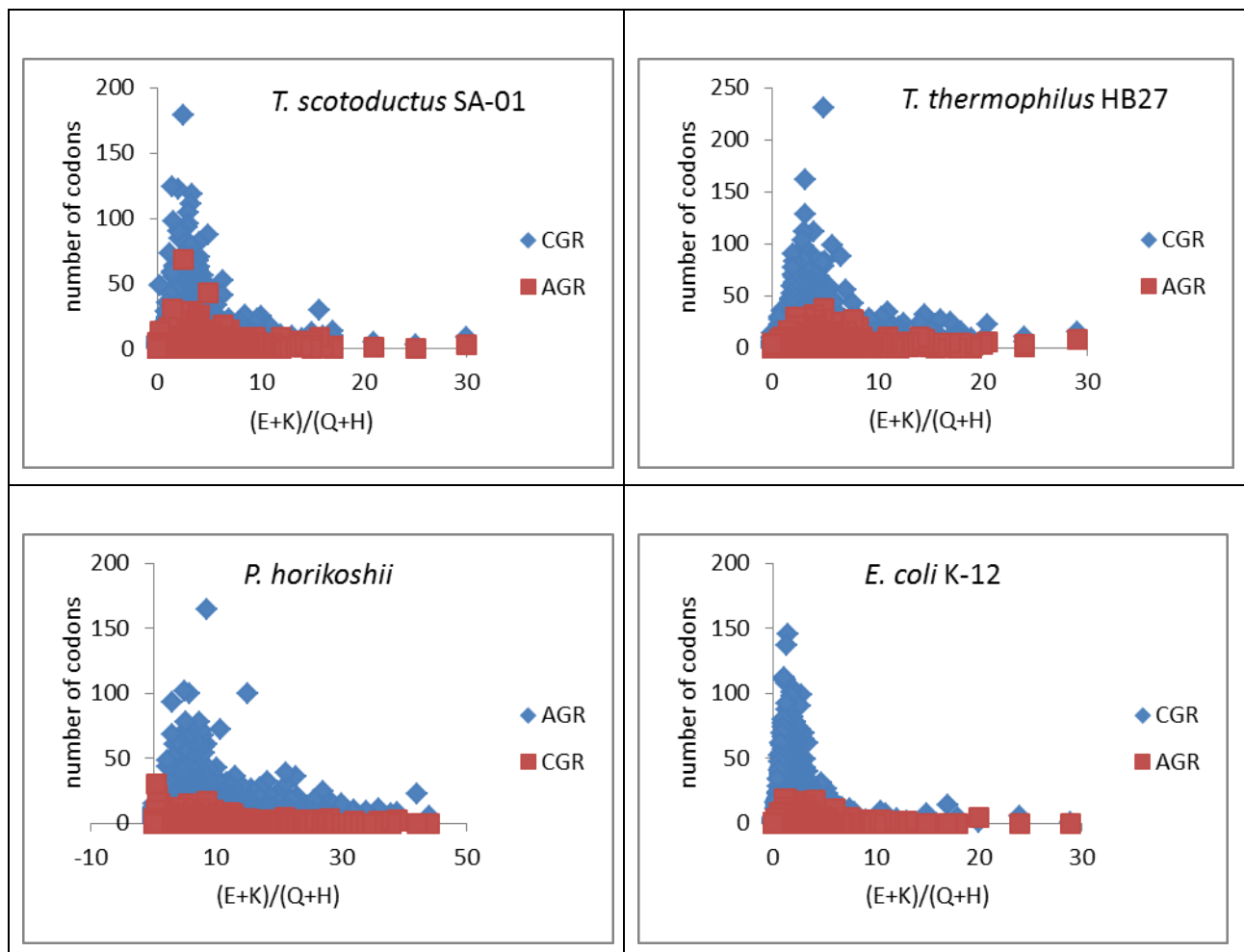


Figure 4.3.2: Codon usage in mesophiles, thermophiles and hyperthermophiles against $(E+K)/(Q+H)$ ratio

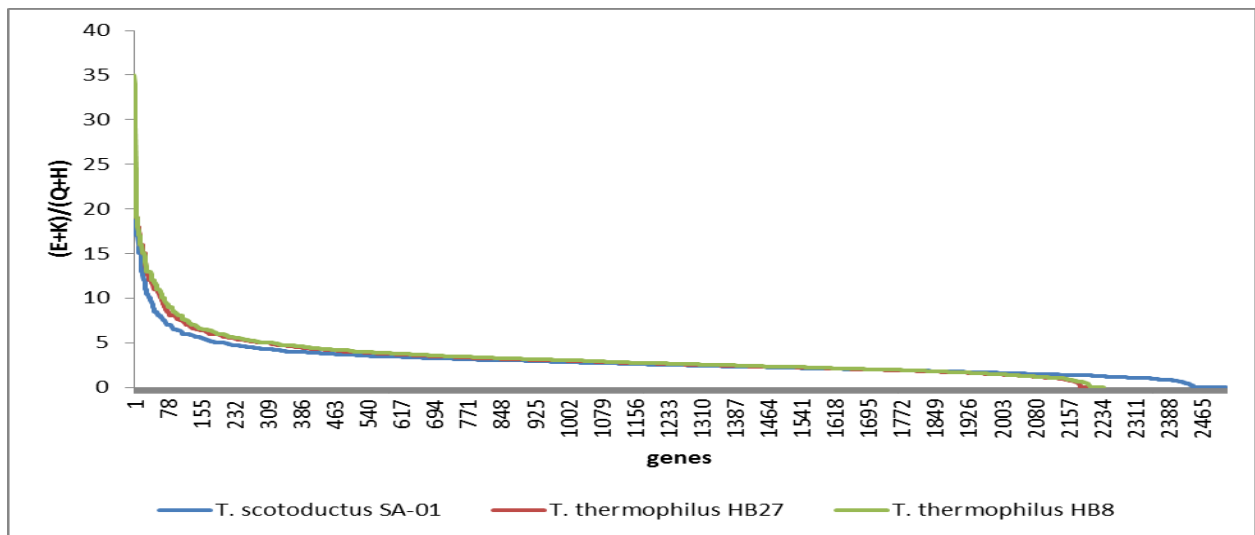


Figure 4.3.3: Thermostability of genes measured using the $(E+K)/(Q+H)$ ratio

4.3.2 Thermostability analysis using minimum folding energy

A python script was implemented to efficiently execute the UNAFold algorithm on all coding sequences. The energy values were translated to positive by negation and normalised by the sequence length. Minimum folding energies for top 500 coding sequences in seven genomes were sorted from largest to smallest and plotted on a graph shown in Figure 4.3.4. These were averages of minimum folding energy values for groups of coding sequences comprising 0.5% of total number of genes in a genome with values sorted from largest to smallest. The 0.5% grouping was used for clear presentation for all coding sequences in genomes.

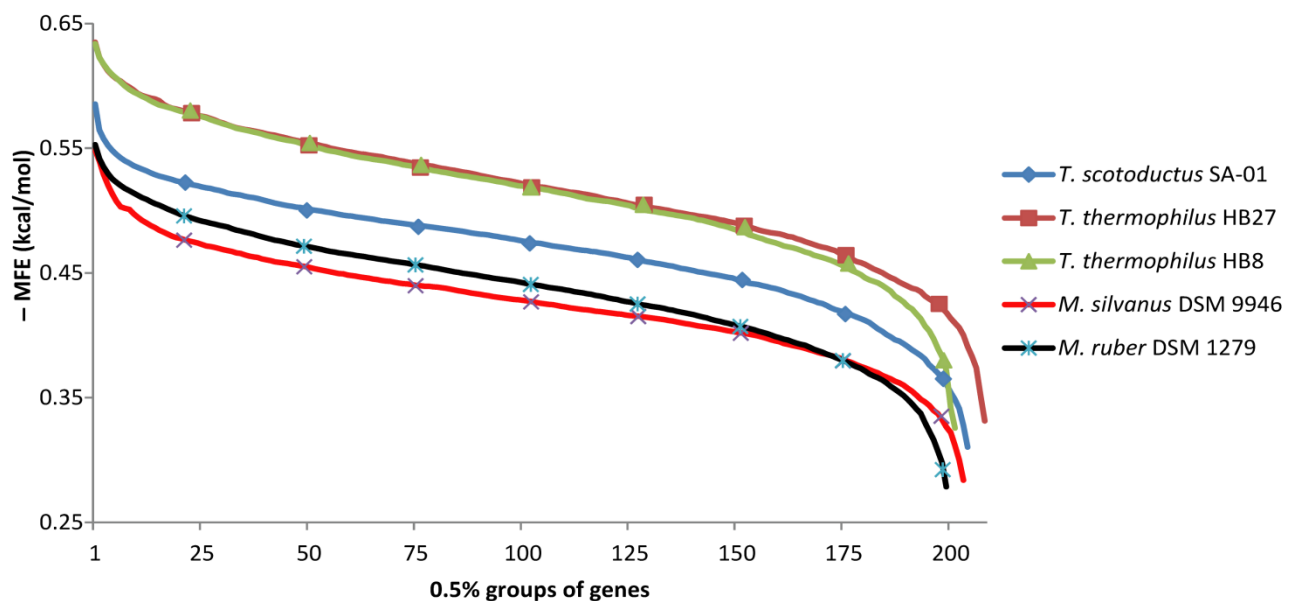


Figure 4.3.4: minimum folding energy for 0.5% groups of genes

Observations from the graph for minimum folding energy values revealed a pattern in their distribution among different genomes. The grouping of minimum folding energy values was consistent with their phylogenetic classification based on 16S rRNA sequences (Figure 2.4.1) and their levels of thermostability as determined by environmental optimum growth temperatures. Thermophilic bacteria (*T. thermophilus* HB8 and *T. thermophilus* HB27 and *T. scotoductus* SA-01) are the top three in the graph. Moderately thermophilic, *M. silvanus* DSM 9946 and *M. ruber* DSM 1279 were ranked next.

On the level of RNA and DNA sequences, this adaptation was achieved by accumulation of base stacking energy rich oligomers in coding sequences. The increase in mRNA thermostability correlated to the general increase in GC-content of analysed genomes from 62-63% in *M. ruber* DSM 1279 and *M. silvanus* DSM 9946 to 69% in *T. thermophilus* strains. This observation however, could not be explained by random substitution of AT pairs of nucleotides to energy rich GC pairs. Comparison of tetranucleotide patterns of concatenated coding sequences of *T. scotoductus* SA-01 against those of *T. thermophilus* HB27 showed a relative increase of energy rich and decrease of energy poor oligomers in *T. thermophilus*. For example, in *T. thermophilus* strains the frequency of GAGG (-33.11 kcal/mol) increased to an average of 194 oligomers per 100 kbp from 126 oligos per 100 kb in *T. scotoductus* SA-01, and the frequency of TTTC (-27.33 kcal/mol) respectively decreased from 38 to 18 oligos per 100 kbp. Similar changes in frequencies of several other oligonucleotides were also observed; however, the choice between preferable and discarded oligonucleotides was selective. For example, frequencies of TTTG and TTGT motifs, and GCGC remained the same in coding sequences of both bacteria, but the frequency of the latter doubled in non-coding sequences of *T. thermophilus* HB27.

Minimum folding energy was used to analyse possible factors that enhance thermostability in *Thermus* species. The measure was preferred because it took into consideration composition, ordering of bases and actual folding of RNA structures. It was also independent of the environmental optimum growth temperature of an organism. As this computation was performed at RNA level, it was theorised that these differences were reflected in protein sequences at amino acid level suggesting positive selection for organisms observed in their genomes to adapt different environmental conditions. To confirm this, orthologous sequences were identified between *T. scotoductus* SA-01 and *T. thermophilus* HB27 with largest minimum folding energy difference.

Interestingly, the top 100 thermostable mRNAs in *T. scotoductus* SA-01 were not the same as the top 100 ones in *T. thermophilus* strains as shown in Appendix 6, Table 6.4.1 and Table 6.4.2. They overlapped by 35% and 30% for *T. thermophilus* HB27 and *T. thermophilus*

HB8, respectively. Thus, some mRNAs acquired more thermostability than others. In an attempt to identify the mechanisms of enhancing thermostability in *Thermus*, correlation was computed between differences in minimum folding energy values calculated for orthologous proteins in *T. thermophilus* HB27 and *T. thermophilus* HB8, and *T. scotoductus* SA-01, and the rates of nucleotide and amino acid substitutions. This difference was examined for correlation with mutational parameters such as: identity, difference, specificity, mutations and sense mutations. Identities are number of amino acids exact matches; differences are given by the formula $(1.0 - \text{identities}) / \text{alignment length}$; specificity is the difference between positives and identities divided by identities; the rate of sense mutations is equal to $(100 \times (\text{indels} + \text{mutations affecting amino acids}) / \text{all mutations})$. The increase in minimum folding energy values in *T. thermophilus* HB27 and *T. thermophilus* HB8 calculated for predicted mRNA secondary structures correlated with both the rate of nucleotide substitutions ($r = 0.38$); and amino acid substitutions ($r = 0.30$) in encoded proteins (Appendix 6, Table 6.4.1 and Table 6.4.2.). These correlations were statistically significant as confirmed by a *t*-test. It was assumed that the correlation between changes in minimum folding energy and frequencies of nucleotide and amino acid substitutions between orthologous protein coding genes implied a parallel adaptation of mRNA and proteins to higher temperatures. A statistically reliable correlation of 0.18 was also found between changes in minimum folding energy and dN/dS non-synonymous/synonymous nucleotide substitution rate ratios (Appendix 6, Table 6.4.1 and Table 6.4.2.). An increased frequency of dN substitutions over dS indicated a positive selection. Thus it was concluded that significant changes in minimum folding energy and possible higher thermostability of mRNA molecules and encoded proteins were under positive evolutionary selection in these micro-organisms.

These findings supported the hypothesis that the adaptation of mRNA to higher temperatures was associated with the increased thermostability of protein molecules. Based on these findings, there was need for further analysis of protein sequences and structures. First, composition and substitution of amino acids in relation to thermostability was investigated. Differences in adaptation for different organisms could not necessarily correlate to individual distribution of amino acids but possibly their properties as well. Hence, the composition and substitution of amino acid residues were computed from which properties were derived and analysed.

4.3.3 Composition of amino acid properties and residues

In a comparative analysis of the genomes of *T. thermophilus* HB27 and *T. scotoductus* SA-01, 1526 orthologous sequences were identified. Out of which 500 sequences with largest differences

in minimum folding energy were analysed for composition of amino acid properties and residues. These 500 orthologs represented a quarter of the total number of orthologs which were representative enough for a generalised conclusion. Sequences from *T. thermophilus* HB27 were more thermophilic than those from *T. scotoductus* SA-01 as measured by minimum folding energy and previously shown in Figure 4.3.4. Analysis of closely related *Thermus* species was aimed at avoiding other factors that could exist between distantly related species. Differences in distantly related organisms could not necessarily be due to differences in thermostability.

Figure 4.3.5 shows differences in composition of amino acid residues between highly thermophilic sequences in *T. thermophilus* HB27 and less thermophilic sequences in *T. scotoductus* SA-01. The sequences were considered to be independent samples since they were from two different genomes. The working hypothesis was that there was no difference in the distribution of amino acid residues between high and less thermophilic sequences in *T. thermophilus* HB27 and *T. scotoductus* SA-01 respectively.

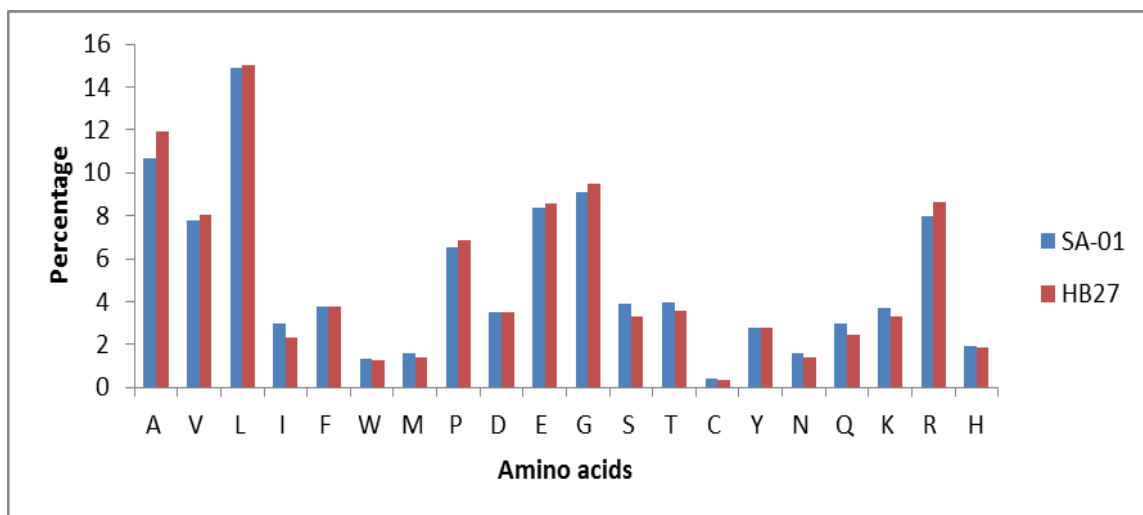


Figure 4.3.5: Composition of amino acid residues between *T. thermophilus* HB27 sequences and *T. scotoductus* SA-01 sequences.

The distribution of each amino acid residues was independently analyzed between the two species to determine if their difference was statistically significant. The hypothesis of preferable use of amino acids in thermostable proteins was rejected at a significant level of 0.05 (Appendix 6, Table 6.5.1). High occurrence of Ile, Ser, Thr, Asn, Gln and Lys was observed in less thermophilic *T. scotoductus* SA-01 sequences while occurrence of Pro, Gly, Arg and Ala were higher in *T. thermophilus* HB27.

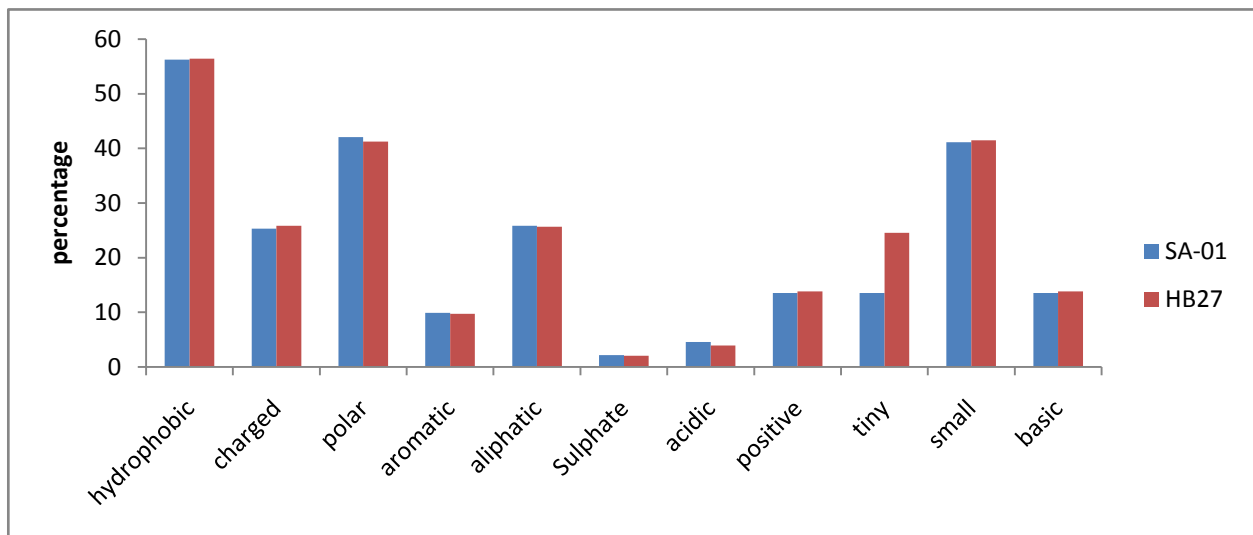


Figure 4.3.6: Differences in amino acid property composition

The aim was to elucidate amino acid properties that influenced differences in thermostability between orthologous sequences between closely related species. Figure 4.3.6 shows the amino acid property composition between orthologous sequences of *T. scotoductus* SA-01 and *T. thermophilus* HB27. The same 500 orthologs with largest minimum folding energy difference were analysed. Sequences from *T. thermophilus* HB27 were highly thermophilic while those from *T. scotoductus* SA-01 were relatively less thermophilic. To ensure that differences were statistically significant, Shapiro-Wilk normality test was made on the distribution of each amino acid property. The parametric *t-test* was applied on normally distributed properties and nonparametric Wilcoxon *t-test* was tested on properties that were not normally distributed. Differences were statistically significant at 95% level of confidence in distribution of polar, sulfur containing, acidic, positively charged, tiny, small and basic amino acids (Appendix 6, Table 6.5.1). Polar, sulfur containing and acidic, were dominant in *T. scotoductus* SA-01, while *T. thermophilus* HB27 had high occurrence of non-polar, positively charged, tiny, small and basic residues. These results agreed with frequencies of amino acid substitutions in Table 4.3.1.

Table 4.3.1 shows differences between direct substitutions from amino acids shown in rows; columns are pairs of orthologous proteins of *T. thermophilus* HB27 and *T. scotoductus* SA-01. Positive values in the Table 4.3.1 implied that the amino acids in columns were accumulated in proteins of *T. thermophilus* HB27. An overall increase or decrease of the amino acids in the sampled proteins of *T. thermophilus* HB27 was indicated in the difference row. The *Increased* and *Decreased* rows show the overall increase or decrease of amino acids in column titles. The expectation was that the absolute values of amino acid substitutions should be proportional to the

frequencies of these amino acids in the sampled proteins. Frequency values for each amino acid are shown in the AMC column.

Table 4.3.1: Pair wise substitutions of amino acids in orthologous proteins of thermotolerant *T. scotoductus* SA-01 and thermophilic *T. thermophilus* HB27.

	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr	AMC
Difference	1836	-2	63	128	-184	415	-57	-986	-795	143	-185	-263	234	-823	1297	-816	-560	690	-60	-75	
Increased								Val	Arg			Asp		Arg,Glu		Ala	Ala			His	
Decreased	Ser,Thr		Asn	Gln			Tyr							Gln,Lys			Ile				
Ala	0																				32594
Cys	0	0																			936
Asp	20	0	0																		9433
Glu	290	-1	13	0																	25858
Phe	23	-1	0	-3	0																10072
Gly	10	-1	-47	-39	-15	0															25751
His	22	0	8	20	-13	8	0														5125
Ile	46	0	0	9	9	1	4	0													6434
Lys	71	0	11	110	1	47	-5	-1	0												9729
Leu	83	4	5	6	-112	19	-14	-227	-9	0											43451
Met	40	1	0	20	4	9	0	2	0	44	0										3299
Asn	52	-1	56	25	4	40	10	0	3	5	-1	0									3745
Pro	19	0	-5	22	-1	-6	-3	-4	-30	-32	-3	-7	0								18184
Gln	108	1	12	249	0	35	16	0	16	46	2	-2	13	0							7155
Arg	47	-5	-4	-92	-16	19	-110	-11	-541	-71	-13	-41	26	-304	0						24379
Ser	416	2	21	51	5	114	12	-6	7	-1	-4	-26	89	-5	92	0					9364
Thr	349	1	14	41	-5	33	0	-9	2	13	4	11	30	-6	42	-34	0				9772
Val	235	-1	0	2	-22	2	1	-656	-7	-110	-52	-5	20	-11	10	-13	-79	0			22490
Trp	-3	1	1	11	16	3	7	-2	0	0	2	0	-2	1	19	1	3	-2	0		3126
Tyr	8	-2	-2	-2	-20	-1	70	-3	-2	4	0	0	8	-2	18	-3	2	6	-4	0	7595

Predominant accumulation of non-polar alanine and positively charged arginine was observed in thermostable proteins of *T. thermophilus* HB27. The accumulation of non-polar amino acids in proteins of thermophiles was reported to increase rigidity and hydrophobicity of proteins (Chakravarty & Varadarajan, 2002). As it was reported by these authors, alanine residues with methyl groups occurred with lower frequency in exposed state but with a higher frequency in well-buried states in thermophilic proteins. Alanine is believed to be the best helix forming residue in thermophiles (Pack & Yoo, 2004; Kumar *et al.*, 2000). Alanine residues are mostly substituted by serine and threonine in the orthologous *T. scotoductus* SA-01 proteins. Replacement of serine by alanine residues was reported as a factor to increase the thermostability of proteins in *Methanococcus*, *Bacillus* (Haney *et al.*, 1999; McDonald, 2010) and *Corynebacterium* (Nishio *et al.*, 2003).

Serine residues tend to impair hydrophobic interactions between beta-strands, whereas alanine is known to be effective in bridging strands (Nishio *et al.*, 2003). Substitution of serine residues by alanine in thermophiles most likely occurs in a series of steps with intermediate glycine residues (Argos *et al.*, 1979; Garg & Johri, 1979) consistent with the observed substitution increase of serine to glycine (Table 4.3.1). Accumulation of threonine in thermostable proteins was also observed (Garg & Johri, 1979) in contradiction to these results of predominant substitution with alanine which is probably specific to *Thermus* species. However, it was known that threonine, as well as serine interacts with water molecules, which increase instability of proteins at higher temperatures (Mattos, 2002). Arginine and glutamate residues, which are accumulated in *T. thermophilus* HB27 proteins (Table 4.3.1), stabilise the exposed

structure of thermophilic proteins. Arginine has a reduced chemical reactivity and high tendency to participate in salt-bridge interactions (Kumar *et al.*, 2000; Chakravarty & Varadarajan, 2002; Das *et al.*, 2006). Several studies have reported high incidence of salt-bridges and ion-pairing in thermophilic proteins (Vogt *et al.*, 1997; Scandurra *et al.*, 1998; Yip *et al.*, 1995).

In thermostable proteins of *Thermus* species, arginine residues predominantly substituted polar glutamine and lysine. Polar residues were much lower in thermophilic proteins. Asparagine and glutamine residues undergo deamination at high temperatures (Cantanzano *et al.*, 1997). Avoidance of lysine and replacement of isoleucine by valine were reported in proteins of some thermophiles, but lysine and isoleucine are frequent in proteins of thermophilic *Methanococcus* (McDonald *et al.*, 1999; Fariás & Bonato, 2002), thus the role of these amino acids in thermostability is not clear. The frequency of histidine was generally reduced in *T. thermophilus* HB27, but in comparison with tyrosine, the former amino acid was preferred in proteins of thermophilic *T. thermophilus* HB27 than in their orthologs from thermotolerant *T. scotoductus* SA-01. It contradicts the previous publications that His/Tyr sites have histidine in mesophiles and tyrosine in thermophiles (McDonald, 2010). Once more this demonstrates the taxonomic specificity of the preferable accumulation of amino acids in *Thermus* proteins. This specificity may be associated with the fact that the high temperature of the environment is not the only selective factor that bacteria have to adapt. Particularly, as it was discussed before, proteins of *Thermus* species are characterized with the increased resistance to many other adverse conditions making them more suitable for biotechnology application (Niehaus *et al.*, 1999). To elucidate possible adaptive reasons of *Thermus* organisms, predominant amino acid substitutions in sequence alignments against calculated 3D structures for selected proteins were mapped.

4.3.4 Protein structure homology modelling

Protein structures were modelled and analysed to identify structural differences that correlated to thermostability. Mutations identified in sequence analysis were mapped into the structure to locate their positions and changes caused which could be associated with thermostability. Analysis was performed in *T. scotoductus* SA-01 and *T. thermophilus* HB27 which are closely related organisms selected to eliminate other evolutionary factors found in distantly related species. Out of all orthologs that met the selection criteria five PDB template structures were identified from orthologs with highest matches with protein structures in PDB as shown in Table 4.3.2.

Among them there were two proteins of *T. thermophilus* HB27 (TTC1891 and TTC1937) which were already crystallized. These were used to model the structures of *T. scotoductus* SA-01

proteins TSC_C20070 and TSC_C00450. The template 2DP9 was used to build models for two ASCH domain containing proteins TSC_C20070 (*T. scotoductus* SA-01) and TTC1891 (*T. thermophilus* HB27) shown in Figure 4.3.7 A. PyMOL was used to locate mutations and visualise changes within structures. Deletion of a short helix in *T. thermophilus* HB27 was observed, with proline in *T. scotoductus* SA-01 mutated to a polar glutamate residue in *T. thermophilus* HB27 (indicated in Figure 4.3.7 A by an arrow *a*). In the same orthologous sequences, a relatively shorter beta-sheet was noticed in the structure of the thermophilic protein of *T. thermophilus* HB27 with no mutational differences observed (arrow *b*). Protein alignment analysis of TTC1891 against TSC_C20070 revealed one Asn =>Asp, two Gln =>Glu, two Ile =>Val and three Lys =>Arg substitutions located predominantly in surface oriented loops and beta-sheets.

Table 4.3.2: Matched orthologs for homology modelling

Templates (PDB ID)	Annotation for modelled orthologous sequences	MFE (kcal/mol)	Sequence identity (%)	Coverage (%)	Coverage Range	E-value	DOPE Z score
2DP9	TSC_C20070 ASCH domain superfamily protein (<i>T. scotoductus</i> SA-01)	-0.39	82	96.77	3-124	1.4E-49	-1.56
	TTC1891 hypothetical protein (<i>T. thermophilus</i> HB27)	-0.50	100	82.55	3-124	9.9E-41	-0.65
2EBJ	TSC_C13250 peptidase (<i>T. scotoductus</i> SA-01)	-0.36	69	99.48	1-192	2.7E-62	-2.42
	TTC0531 peptidase (<i>T. thermophilus</i> HB27)	-0.49	99	99.48	1-192	5E-63	-2.42
2CWY	TSC_C00450 conserved hypothetical protein (<i>T. scotoductus</i> SA-01)	-0.44	60	93.62	5-94	5.9E-35	-2.19
	TTC1937 hypothetical protein (<i>T. thermophilus</i> HB27)	-0.54	100	100	1-94	3.7E-38	-2.78
2FK5	TSC_C04250 fuculose-1-phosphate aldolase (<i>T. scotoductus</i> SA-01)	-0.48	83	94.5	1-190	1.4E-51	-1.78
	TTC1459 L-fuculose phosphate aldolase (<i>T. thermophilus</i> HB27)	-0.57	98	99.5	1-200	1E-54	-1.73
1V8D	TSC_C14180 conserved hypothetical protein (<i>T. scotoductus</i> SA-01)	-0.46	80	82.55	40-235	5E-104	-0.84
	TTC0214 transcriptional regulator (<i>T. scotoductus</i> SA-01)	-0.55	95	82.55	40-235	5E-107	-1.14

The template 2EBJ was used to construct 3D models of two peptidases TSC_C13250 (*T. scotoductus* SA-01) and TTC0531 (*T. thermophilus* HB27) shown in Figure 4.3.7 A-B. No differences were observed in the structures (Figure 4.3.7 B). Protein alignment analysis of TTC0531 against TSC_C13250 revealed one Ser =>Ala, one Thr =>Ala, two Ile =>Val and one Lys =>Arg substitutions located in surface oriented loops and domains. The template 2CWY was used to build 3D models of two hypothetical proteins TSC_C00450 (*T. scotoductus* SA-01) and TTC1937 (*T. thermophilus* HB27) shown in Fig. 4.3.8 C. There were no differences observed in the structures. Protein alignment analysis of TTC1937 against TSC_C00450 revealed three Lys =>Arg and one Glu =>Arg substitutions located in surface oriented alpha-helices.

The template 2FK5 was used to create 3D structures of two fuculose phosphate aldolases TSC_C04250 (*T. scotoductus* SA-01) and TTC1459 (*T. thermophilus* HB27) shown in Figure 4.3.7 D. TSC_C04250 contains one additional surface alpha-helix that was missing in the structure of more thermostable TTC1459 (arrow *c* in Figure 4.3.7 D) although there were no mutational differences observed within this region. Protein alignment analysis of TTC1459 against TSC_C04250 revealed one Ser =>Ala, one Tyr =>His, two Ile =>Val and two Lys =>Arg substitutions located mostly in surface oriented beta-sheets.

The template 1V8D was used for modelling 3D structures of the conserved hypothetical protein TSC_C14180 from *T. scotoductus* SA-01 and the orthologous transcriptional regulator TTC0214 from *T. thermophilus* HB27 shown in Figure 4.3.7 E. TSC_C14180 contains one additional surface alpha-helix that was missing in the structure of more thermostable TTC0214 (arrow *d*) without any mutational differences observed at sequence level. Several mutations have been observed located on the surface and loops of the structure. The mutations were: one three Thr =>Ala, four Ile =>Val, four Lys =>Arg and one Glu =>Arg. In order to determine the stabilising effect of each of these mutations, FoldX plugin for Yasara program version 1.4.21 (Schymkowitz *et al.*, 2005; van Durme *et al.*, 2011) was used to calculate the energy difference between the structure and the wild type. The original structures that matched with *T. thermophilus* HB27 were analysed. As shown in Table 4.3.3, these had high sequence identity, 100% for 2DP9 and 2CWY; 99% for 2EBJ; 98% for 2FK5 and 95% for 1V8D.

However, in order to be precise only structures with sequence identity greater than 95% were analysed in the stability change analysis leaving out structure 1V8D. Since FoldX has a ± 0.5 kcal/mol error margin, values above the threshold were considered significant. Only predominant mutations identified in this study as shown in Table 4.3.1 were analysed for the stabilising effect on protein structures. Negative values indicated a greater energy, implying stabilisation, while positive values meant destabilising effect.

Table 4.3.3: Impacts of predominant amino acid substitutions on protein stability predicted by FoldX.

PDB ID	Mutation SA-01 => HB27	Property Change	Position	Location in 3D	Energy change (kcal/mol)
2DP9	Lys => Arg	Conserved	3	S, L	-0.61*
	Lys => Arg	Conserved	29	S, L	+0.03
	Ile => Val	Conserved	41	B, BS	+0.61*
	Asn => Asp	To charged	51	S, BS	-0.09
	Gln => Glu	To charged	55	S, L	0.00
	Gln => Glu	To charged	56	S, BS	+0.10
	Ile => Val	Conserved	104	B, L	+0.52*
	Lys => Arg	Conserved	106	S, L	-0.48
2EBJ	Thr => Ala	To non-polar	20	S, H	-1.50*
	Ile => Val	Conserved	38	S, BS	+0.48
	Ser => Ala	To non-polar	43	S, L	-0.86*
	Ile => Val	Conserved	80	S, BS	-0.49
	Lys => Arg	Conserved	129	S, BS	-1.20
2CWY	Glu => Arg	To charged	9	S,H/L	-0.82*
	Lys => Arg	Conserved	49	S,H	-0.32
	Lys => Arg	Conserved	59	S, H	+0.19
	Lys => Arg	Conserved	64	S,H	+0.31
2FK5	Lys => Arg	Conserved	4	S, H	+0.27
	Ser => Ala	To non-polar	7	B,H	-0.75*
	Ile => Val	Conserved	73	S, H	-0.14
	Tyr => His	To charged	113	S, H	+0.30
	Lys => Arg	Conserved	141	S,H	+0.25
	Ile => Val	Conserved	156	S,BS	+0.38
1V8D	Lys => Arg	Conserved	6	L,B	N/A
	Lys => Arg	Conserved	10	H,S	N/A
	Ile => Val	Conserved	12	H,B	N/A
	Lys => Arg	Conserved	40	L,S	N/A
	Lys => Arg	Conserved	44	H,S	N/A
	Ile => Val	Conserved	53	H,B	N/A
	Ile => Val	Conserved	66	L,S	N/A
	Ile => Val	Conserved	68	BS,B	N/A
	Thr => Ala	To non-polar	89	H,S	N/A
	Thr => Ala	To non-polar	96	L,S	N/A
	Thr => Ala	To non-polar	106	B, H	N/A
	Glu => Arg	To charged	181	H,S	N/A

The location of substitutions was marked as S – surface; BS - beta sheet; H - helices and L – in loops. Values of energy change above the FoldX error margin of ± 0.5 kcal/mol are marked by asterisks. FoldX calculations were not applicable for the template 1V8D due to its low identity (95%) with the studied proteins.

As shown in Table 4.3.3, the effect of individual amino acid substitutions varied and often was below the FoldX sensitivity. The cumulative effect of studied substitutions was either neutral for templates 2DP9 and 2FK5, or showed weak protein stabilization in the templates 2EBJ and 2CWY. It can be concluded that protein stabilisation due to increase in folding energy cannot be the only factor driving protein adaptation to higher temperatures. Redressing of surface amino acids from polar to non-polar; uncharged to charged and possibly adjusting of protein-protein interactions are other means of the high temperature adaptation. In terms of affecting the protein

folding energy the same substitutions may have different effect depending on its location and neighbouring atoms on the protein structure.

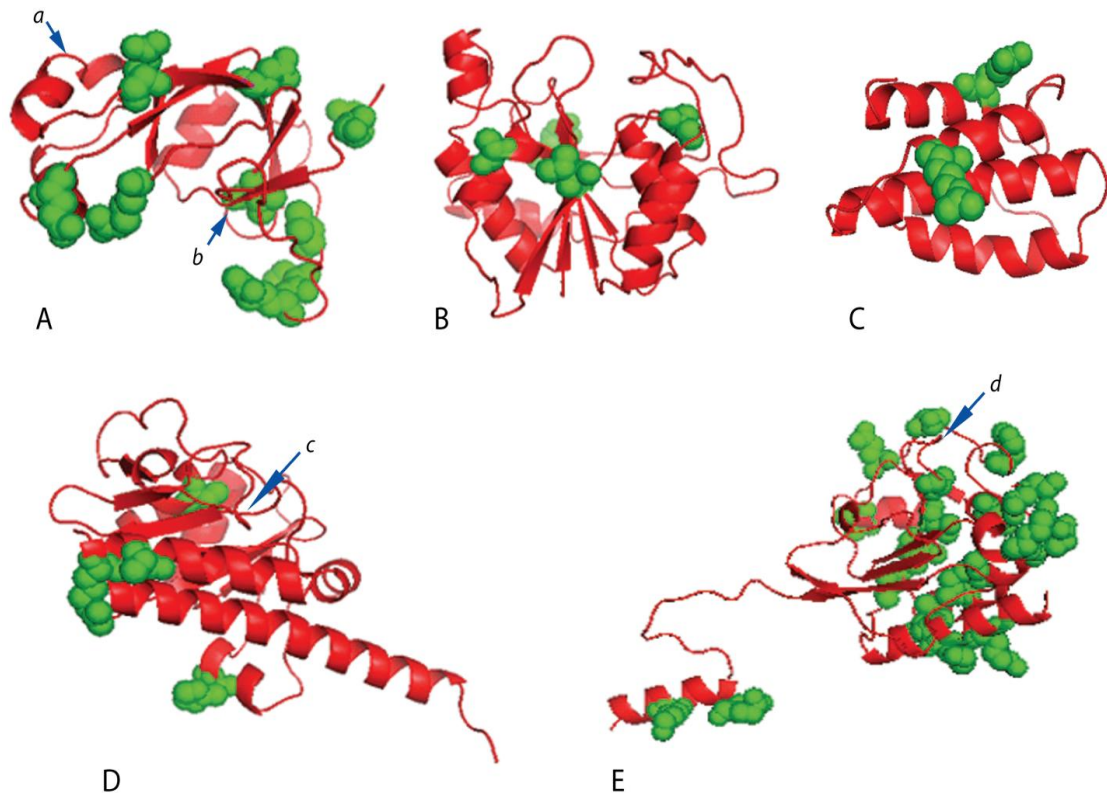


Figure 4.3.7: Comparison of thermophilic and mesophilic protein structures

Protein structural models based on templates A) 2DP9; B) 2EBJ; C) 2CWY; D) 2FK5; and E) 1V8D. Positions of minor structural changes between the orthologous proteins of *T. scotoductus*SA-01 and *T. thermophilus*HB27 are depicted by arrows *a*, *b*, *c* and *d*.

4.4 Conclusion

Analysis of proteins sequences and structures was aimed at determining factors that enhance thermostability in *Thermus* species. Contrary to prior studies which analysed thermostability enhancing factors between distantly related species such as mesophiles and thermophiles, this work analysed closely related thermophilic species of the family Thermaceae specifically the thermotolerant *T. scotoductus* SA-01 against thermophilic *T. thermophilus* HB27. This was to eliminate other evolutionary factors with no effect on thermostability. The UNAFold algorithm was extended in this work to efficiently handle large data sets to compute MFEs for coding sequences in entire genomes. The increase in minimum folding energy in *T. thermophilus* correlated with the general increase in the genomic GC-content. Minimum folding energy may be a useful measure of distinguishing thermostable proteins of complete genomes of closely related

organisms. However, it may not be applicable for comparison in distantly related organisms with diverse GC-content. Minimum folding energy values calculated for AT-rich sequences would be smaller, which could not necessarily imply that mRNAs are more sensitive to high temperatures, as they may have different mechanisms of mRNA stabilisation (Erickson & Gross, 1989; Takayama & Kjelleberg, 2000; Garrett & Rosenthal, 2012; Alexandre & Olivier, 2012).

Adjustment of proteins for higher thermostability was found to be highly affected by amino acid substitutions. A pattern of predominant amino acid substitutions was shown in Table 4.3.1. It was observed in this work that Fariás-Bonato ratio (Glu+Lys)/(Gln+His) recommended for classifying thermophilic and mesophilic proteins (Fariás & Bonato, 2002) was not applicable for Thermaceae because of unexpected decrease of lysine residues and ulterior differences in frequencies of glutamate and histidine (Table 4.3.1). This observation indicates that there are different mechanisms of evolutionary adaptation of proteins to stresses exploited in different bacterial taxa. Fariás-Bonato ratio may be suitable for distantly related organisms but may not for closely related species.

Orthologous coding sequences from *T. scotoductus* SA-01 and *T. thermophilus* HB27 with largest difference in minimum folding energy values were analysed. Localization of predominant amino acid substitutions highlighted in Table 4.3.3 was studied in 3D structural models of five *T. thermophilus* HB27 and *T. scotoductus* SA-01 proteins, for which appropriate templates were modelled. A general trend of localization of substitutions was observed in protein surfaces at the ends of conserved domains and in loops.

Alteration of loops was observed in several studies to affect protein stabilisation (Nagi *et al.*, 1999; Sanchez, 2008; Russell *et al.*, 1997). Interestingly, in all five structures, the predominant substitutions did not change the overall structure of proteins. The observed gain of surface exposed tiny, small and charged amino acid residues in *T. thermophilus* HB27 contributed to the stabilization of proteins at high temperatures by reduced chemical reactivity and high tendency to participate in salt-bridge interactions (Kumar *et al.*, 2000; Chakravarty and Varadarajan, 2002; Nishio *et al.*, 2003; Das *et al.*, 2006), an increased rigidity and hydrophobicity (Chakravarty & Varadarajan, 2002; Mattos, 2002) and stabilization of helices (Pack & Yoo, 2004; Kumar *et al.*, 2000).

Amino acid substitutions contributed to protein stabilization by increasing the overall protein folding energy. FoldX program was used to estimate the energy change of predominant substitutions (Table 4.3.3). A weak energy increase was reported for the templates 2EBJ and 2CWY, but the overall difference in the templates 2DP9 and 2FK5 was statistically unreliable. It may be concluded that the adaptation of proteins to a high temperature environments was driven

by multiple factors acting in synergistic manner. Structural changes were detected in three out of five structures (Figure 4.3.7). In all cases the conserved domains disappeared or became shorter in proteins of *T. thermophilus* HB27 consistent with published report that shortening of loops and domains contribute to compactness of proteins, which is essential in achieving protein thermostability (Russell *et al.*, 1997). Structural changes were associated with Pro=>Glu substitution in TSC_C20070 in comparison to TTC1891 proteins (Figure 4.3.7, substitution *a*). The energy contribution of each substitution differed depending on its location and neighbours in the structure. The overall substitution analysis showed that stabilizing effect (negative energy values) of mutations, were much higher; mostly above the error threshold of ± 0.5 kcal/mol for FoldX as compared to destabilizing effect (positive energy values) on protein structures. This showed that the abundance of such mutations could indeed lead to a more stable protein structure as previously observed

Chapter 5

Concluding Discussion

5.1 Introduction

Thermophilic bacteria of genus *Thermus* were the main focus of this study. These bacteria inhabit both man-made and natural higher temperature environments such as hot springs, deep mines and domestic hot water (Oshima & Imahori, 1974; Munster *et al.*, 1986; Brock & Freeze, 1969). Their ability to live in such high temperature environments has fascinated biological scientist for many years (Lasa & Berenguer, 1993). Enzymes from these bacteria are able to function at higher temperatures without denaturing to support biochemical processes and life. Their metabolic system has evolved to function in extreme conditions to sustain life. Their enzymes are resistant to extreme conditions such as acidity; alkaline conditions and salt concentration (Niehaus *et al.*, 1999).

The genome of *Thermus* bacteria is believed to be highly competent due to high levels of natural transformation, a major mechanism for horizontal gene transfer through which genes are acquired from the environment from dead cells. Natural transformations are considered to be their major survival technique in higher temperature environments. The acquiring and loss of genes through horizontal gene transfer is believed to affect the order and arrangement of genes on the chromosome (Schwarzenlander *et al.*, 2009, Friedrich *et al.*, 2003). It is known however, that functionally related genes such as those acting in similar pathways or substrates are localised in proximity on the chromosome and are usually co-expressed, fused or located in the same operon (Spirin *et al.*, 2006). There is a correlation between the order of genes on the chromosome and the organisation and structure of the metabolic network. This means that any changes in the order of functionally related genes on the chromosome affect the coherence and structure of the metabolic network. Changes in the structure of the metabolic network occur due to pressure of evolutionary forces which modularise the network to perform specialised functions that are necessary for survival in hostile environments (Fani & Fondi, 2009). By comparing structures of different metabolic networks, the evolutionary pressures that micro-organism went through can be observed.

Although the study of *Thermus species* dates back to 1960's when Brock and Freeze (1969) isolated *Thermus aquaticus* from the hot springs of Yellow Stone National Park in USA, and that so far over 50 species have been isolated worldwide, only two strains of *Thermus thermophilus* HB8 and HB27 were completely sequenced before this research. It was until

recently at the commencement of this study when *Thermus scotoductus* SA-01 genome was completely sequenced. The strain was isolated from an old gold mine at Witwatersrand in South Africa at a depth of 3.2 kilometres (Kieft *et al.*, 1999). This study was thus inspired by the study of *Thermus scotoductus* SA-01 genome to further investigate horizontal gene transfer, genome rearrangements, metabolic network organisation and factors that enhance thermostability. The phenomenon of extraordinary *Thermus* genome plasticity and its role in acquiring genes encoding thermostable enzymes, formation of metabolic pathway, regulation and biosynthesis at both transcriptional and translational levels remains unclear. The bacteria are known to have twelve distinct competent genes which code for pilin-like proteins similar to type IV pilus biogenesis, eleven of which were identified and implicated in natural transformation of *T. thermophilus* HB27 (Schwarzenlander *et al.*, 2009; Friedrich *et al.*, 2003). Despite these findings, there have been no studies investigating the extent to which *Thermus* genomes have been rearranged as a result of the natural transformations. Factors that enhance protein thermostability have also not been investigated in *Thermus* species. Mostly, these factors are inferred from other thermophilic bacteria and applied to *Thermus* bacteria. However, studies have come up with conflicting results on factors that enhance protein thermostability in thermophilic bacteria. This has been due to different methods that are used to measure thermostability and a limited amount of data which is not sufficient for reliable statistical analysis. Besides, these factors, thermostability varies between taxa up to the species level, but may also differ from protein to protein (Trivedi *et al.*, 2006). Hence, although the factors that affect thermostability have been determined elsewhere, those factors may not be fully applicable for *Thermus* species and require to be tested.

This study was divided into two parts. The first part of the study investigated genome rearrangements and the coherence of metabolic networks as possibly affected by genome rearrangements. The second part of the study analysed factors that enhance protein thermostability in *Thermus* species. The first part of the study addressed two aims. The first aim was to investigate the levels of genome rearrangements in *Thermus* species. Genome rearrangements were analysed by computing the distribution of functionally related genes on the chromosome and the occurrence of breakpoints that resulted from rearrangements of genes or blocks of genes on the chromosome. Thus, breakpoints were defined as the disruption in the synteny of genes due to translocation, inversions, deletions and insertions. Particular attention was paid as to whether location of breakpoints tends to occur in specific regions (hotspots) on the chromosome and whether they disrupted any existing operons or pathways. Determination of breakpoint positions enabled the analysis of hotspots of genome rearrangements which are chromosomal areas highly prone to excision and insertion of genomic fragments. The

contribution of horizontal gene transfer through natural transformations to rearrangement of genes on the chromosome was investigated, particularly the extent to which they trigger genome rearrangements from both physical (acquiring and loss of genes) and functional (functions of acquired genes) points of view. Further analysis was conducted on horizontal gene transfer to estimate the donor-recipient analysis of genomic islands and their possible age. This analysis attempted to establish the movement of genomic islands and their source organisms from which they transferred to *Thermus* species in order to identify possible origins of their adaptative traits. It also provided an indication of the ability for *Thermus* species to acquire foreign genes which pointed to their level of competence in comparison to other bacteria species.

Results of distribution of genes revealed that generally, functionally related genes were in proximity in all chromosomes for mesophilic, moderately thermophilic and thermophilic bacteria. More functionally related genes were found in proximity within the distance category of 0-1000 (Figure 2.6.1). There was fewer occurrence of functionally related genes was observed within largest distance category of 100,000 - 1,000,000 in agreement with hypotheses in prior studies. The distribution of functionally related genes between thermophilic and mesophilic bacteria was expected to be significantly different due to high levels of natural transformations in thermophiles. The earlier assumption based on prior work was that natural transformations would trigger rearrangements of genes on the chromosome. However, this work found that genome rearrangements did not appear to have significantly affected rearrangement of functionally related genes.

Almost double rearrangements were observed in *T. thermophilus* HB8 and *T. thermophilus* HB27 as compared to *T. scotoductus* SA-01 (Figure 2.7.1). The number of rearrangements was consistent with the measure of evolutionary changes as observed in Figure 2.4.1. They also correlated to the environmental temperature for the two *Thermus* species which were between 60 to 65 °C for *T. scotoductus* SA-01; while for *T. thermophilus* HB8 and *T. thermophilus* HB27 temperature ranged from 70 to 85 °C (Balkwill *et al.*, 2004; Opperman & Heerden, 2007; Saiki *et al.*, 1972; Oshima & Imahori, 1974). This finding was consistent with earlier studies which associated high levels of genome rearrangements to levels of thermostability.

The second aim within the first part of the study was to measure the coherence of the metabolic network in order to estimate evolutionary pressures experienced by micro-organisms. As a result of the direct link that exists between organisation of genes on the chromosome and the structure of the metabolic network, it was assumed that rearrangements of genes on the chromosome affect the coherence of the metabolic network. The effect of genome

rearrangements on the organisation of the metabolic network was analysed by clustering metabolic networks. The levels of clustering in *Thermus* species were contrasted with the number of rearrangements in their genomes in attempt to establish a correlation. A clustering index was computed by estimating associations of genes catalysing similar metabolic pathways or metabolites. The level of metabolic network clustering then was contrasted against thermophiles, moderate thermophiles and mesophiles in order to determine whether the clustering correlated with levels of genome rearrangements in *Thermus* bacteria and thermostability in all other bacteria as determined by the optimum growth temperature.

Results of metabolic network clustering revealed a correlation to thermophilic lifestyle as determined by environmental optimum growth temperatures of the organisms shown in Figure 2.7.2. Poorer clustering was observed in *T. thermophilus* HB8 and *T. thermophilus* HB27 as compared to *T. scotoductus* SA-01. Clustering of *M. silvanus* DSM 9946 and *M. ruber* DSM 1279 were similar to that of *T. scotoductus* SA-01. The level of clustering suggested the extent to which the metabolic networks evolved. Metabolic networks for *Thermus* species therefore, seemed to have undergone more evolutionary changes in their organisation which was possibly due to shuffling of genes and disintegration of pathways to suite their thermophilic lifestyle.

The second part of the study analysed thermostability of *Thermus* species, particularly factors that enhanced protein thermostability. This analysis was divided into three sections: firstly, a measure of thermostability applicable to *Thermus* proteins was investigated; secondly, the identified measure was applied to investigate factors that enhance thermostability at protein sequence level; thirdly, identified thermostability enhancing factors were analysed at protein structural level.

The suitability of several measures of protein thermostability in *Thermus* species was analysed. The optimum growth temperature was used in this work as a general estimate for thermostability but obviously it was rather a rough approach which revealed nothing about the thermostability of individual proteins. Melting temperatures of proteins was also not feasible to determine individual thermostable proteins because there is not enough data available for protein melting temperatures from experiments for comprehensive analysis. It was reported that protein thermostability may depend on amino acid content. Nevertheless, Fariás and Bonato ratio $(E+K)/(Q+H)$ which estimates protein thermostability based on amino acid composition was also not applicable for *Thermus* species. Although this ratio correlated with optimum growth temperature of many mesophiles and thermophiles, the values calculated for *T. scotoductus* SA-01 were below the expected ratio for thermophilic organisms. Again, Fariás and Bonato ratio was proposed mostly for prediction of thermostability of micro-organisms by averaging the ratio of all

sequences in the genome of the microorganism. As an approach for predicting thermostable proteins, it was decided to consider minimum folding energy calculated for RNA molecules encoding these proteins. Minimum folding energy was computed by the Unified Nucleic Acid Folding algorithm (UNAFold), an improvement of the mfold algorithm developed by Zuker and Stiegler (1981). Computation of minimum folding energy for RNA sequences showed a distribution pattern which was consistent with the optimum growth temperature of phylogenetically related organisms (Figure 4.3.4). Thus, it can be concluded that this approach may be useful to compare the level of adaptation to higher temperature and probably to other adverse factors among closely related organisms. This approach was also used to compare levels of thermostability of proteins within the same organism to identify those with highest levels of thermostability. An interesting finding was that despite sharing orthologous genes between *T. scotoductus* and *T. thermophilus*, selection of most thermostable proteins as calculated by minimum folding energy were not identical between the two species implying different evolutionary forces for adaptation to thermal environments.

Since Fariás and Bonato ratio was not applicable for measuring thermostability in *Thermus* species, it was interesting to investigate whether there was a correlation between preferred usage of specific amino acids and the level of thermostability of *Thermus* species as estimated by optimum growth temperature. Indeed, such a correlation was revealed with a correlation coefficient for nucleotide and amino acid substitutions 0.30 and 0.38 respectively (Appendix D, Table 6.4.1 and Table 6.4.2).

Minimum folding energy, the identified measure for thermostability in *Thermus* species was applied to investigate factors that enhance protein thermostability. In this study, protein sequences from thermophilic *T. thermophilus* HB27 and their thermotolerant orthologs from *T. scotoductus* SA-01 were compared. Closely related organisms were selected for comparison to eliminate interference of other evolutionary factors. The distribution of each amino acid residue was independently analyzed between the two species to determine if the difference was statistically significant. The parametric *t-test* was applied on normally distributed properties and nonparametric *Wilcoxon t-test* was tested on properties that were not normally distributed. The hypothesis of preferable use of amino acids in thermostable proteins was rejected at a significance level of 0.05 (Appendix 6, Table 6.5.1). Results showed a higher occurrence of Ile, Ser, Thr, Asn, Gln and Lys in less thermophilic *T. scotoductus* SA-01 sequences while occurrence of Pro, Gly, Arg and Ala were higher in *T. thermophilus* HB27. Statistically significant difference at 95% level of confidence was revealed in the property distribution of polar, sulfur containing, acidic, positively charged, small and basic amino acids (Appendix 6,

Table 6.5.1). Polar, sulfur containing and acidic amino acids were dominant in *T. scotoductus* SA-01, while *T. thermophilus* HB27 had higher occurrence of non-polar, positively charged, tiny, small and basic residues. These results agree with the frequencies of amino acid substitutions found in Table 4.3.1. Dominant amino acid substitutions which may contribute to a higher thermostability of proteins were summarised from Table 4.3.1. The major finding was that there was higher dominance of non-polar alanine and positively charged arginine in highly thermophilic sequences. Also a frequent substitution of alanine by serine and threonine, with arginine being substituted by glutamine and lysine in less thermophilic sequences was observed (Table 4.3.1)

Orthologous sequences that met the selection criteria (Table 4.3.2) of the biggest difference between minimum folding energy values were used to search for 3D protein template structures in PDB (Berman *et al.*, 2000) using HHpred online program (Soding *et al.*, 2005). Templates with over 80% identity to the target sequences with a resolution of less than 3Å were used to build 3D structure models using MODELLER v9.11 program (Sali & Blundell, 1993). PyMOL program (DeLano, 2002) was used to determine the locations of mutations as to whether they occur at the surfaces or in buried areas; or whether they were in helices or loops (Table 4.3.3). Most mutations were observed to occur in loops and on the surfaces of protein globules. Shortening of helices was also observed in higher thermophilic proteins however, it was not always possible to associate these structural changes with specific mutations (Figure 4.3.7).

The protein folding energy change contribution of each mutation were analysed in the structures by FoldX algorithm using the FoldX plugin for Yasara program version 1.4.21 (Schymkowitz *et al.*, 2005; Durme *et al.*, 2011). The energy contribution of each substitution differed depending on its location and neighbouring atoms in the structure. The overall substitution analysis revealed that the stabilizing effect (negative energy values) of mutations were much higher mostly above the error margin of ± 0.5 kcal/mol for FoldX as compared to the destabilizing effect (positive energy values) on protein structures. This showed that the abundance of such mutations could indeed lead to more stable protein structures with higher thermostability. This observation allowed a conclusion that substitutions on both nucleotide and amino acid levels contributed to better stabilization of mRNA molecules and proteins of *Thermus* bacteria.

Bibliography

- Alexandre, A., & Olivier, S. (2012) Response to temperature stress in rhizobia. *Crit. Rev. Microbiol.* **10**, doi: 10.3109/1040841X.2012.702097.
- Annison, G. (1992) Commercial enzyme supplementation of wheatbased diets raises ileal glycanase activities and improves apparent metabolisable energy, starch and pentosan digestibilities in broiler chickens. *Animal Feed Science and Technology* **38**, 105-121.
- Anold, R., DiChristina, T., & Hoffmann, M. (1988) Reductive dissolution of Fe(III) oxides by *Pseudomonas* sp 200. *Biotechnology and Bioengineering* **32**, 1081-1096.
- Argos, P., Rossman, M., Grau, U., Zuber, H., Frank, G., & Tratschin, J. (1979) Thermal stability and protein structure. *Biochemistry* **18(25)**, 5698-703.
- Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201.
- Aziz, R., Breitbart, M., & Edwards, R. (2010) Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Research* **38**, 4207-4217.
- Baedecker, M., & Back, W. (1979) Modern marine sediments as a natural analog to the chemically stressed environment of a landfill. *Journal of Hydrology* **43**, 393-414.
- Balkwill, D., Keift, T., Tsukuda, T., Kostandarithes, H., Onstott, T., Macnaughton, S., et al. (2004) Identification of iron-reducing *Thermus* strains as *Thermus scotoductus*. *Extremophiles* **8**, 37-44.
- Becker, P., Abu-Reesh, I., Markossian, S., Antranikian, G., & Márkl. (1997) Determination of the kinetic parameters during continuous cultivation of the lipase-producing thermophile *Bacillus* sp. IHI-91 on olive oil. *Applied Microbiol Biotechnology* **48**, 184-190.
- Beg, Q., Kapoor, M., Mahajan, L., & Hoondal, G. (2001) Microbial xylanases and their industrial applications: a review. *Applied Microbiol Biotechnology* **56**, 326-338.
- Belda, E., Moya, A., & Silva, F. (2005) Genome rearrangement distances and gene order phylogeny in Proteobacteria. *Molecular Biology and Evolution* **22(6)**, 1456-1467.
- Bell, P., Mills, A., & Herman, J. (1987) Biogeochemical Conditions Favoring Magnetite Formation during Anaerobic Iron Reduction. *Applied and Environmental Microbiology* **53(11)**, 2610-2616.
- Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissing, H., et al. (2000) The protein Data Bank. *Nucleic Acids Research* **28(1)**, 235-242.
- Bezuidt, O., Lima-Mendez, G., & Reva, O. N. (2009) SeqWord Gene Island Sniffer: a program to study the lateral genetic exchange among bacteria. *Word Academy of Science, Engineering and Technology* **58**, 1169-1174.
- Bezuidt, O., Pierneef, R., Mncube, K., Lima-Mendez, G., & Reva, O. (2011) Mainstreams of Horizontal Gene Exchange in Enterobacteria: Consideration of Outbreak of Enterohemorrhagic *E. Coli* O104:H4 in Germany in 2011. *PLOS One* **6**, e25702.
- Bostrom, B., Andersen, J., Fleischer, S., & Jansson, M. (1988) Exchange of phosphorus across the sediment-water interface. *Hydrobiologia* **170**, 229-244.

- Boto, L. (2010) Horizontal gene transfer in evolution: facts and challenges. *Proceedings of the Royal. Soci. Biological Sciences* **277**, 819-827.
- Breithaupt, H. (2001) The hunt for living gold. *EMBO reports* **2(11)**, 968-971.
- Britton, K., Baker, P., Borges, M., Engel, P., Pasquo, A., Rice, D., et al. (1995) Insights into thermal stability from a comparison of the glutamate dehydrogenases from *Pyrococcus furiosus* and *Thermococcus litoralis*. *European Journal of Biochemistry* **229**, 688-695.
- Brock, T., & Freeze, H. (1969) *Thermus aquaticus* gen. n. and sp. n., a Non-sporulating Extreme Thermophile. *Journal of Bacteriology* **98(1)**, 289-297.
- Burge, C., Campbell, A., & Karlin, S. (1992) Over- and under-representation of short oligonucleotides in DNA sequences. *Proc. Natl. Acad. Sci. USA* **89**, 1352-1362.
- Canada (1994) Chromium and its compounds. Priority list assessment report. *Chromium and its compounds. Priority list assessment report*. In Canadian environmental protection act
- Cantanzano, F., Capasso, S., & Barome, G. (1997) Thermodynamic analysis of the effect of selective monodeamination at asparagine 67 in ribonucleases A. *Protein Science* **6**, 1682-1693.
- Carter Jr, C., LeFebvre, B., Cammer, S., Tropsha, A., & Edgel, M. (2001) Four-body Potentials Reveal Protein-specific Correlations to Stability Changes Caused by Hydrophobic Core Mutations. *J. Molec. Biol.* **311**, 625-638.
- Chakravarty, S., & Varadarajan, R. (2000) Elucidating of determinants of protein stability through genome sequence analysis. *FEBS Letters* **470**, 65-69.
- Chakravarty, S., & Varadarajan, R. (2002) Elucidation of Factors Responsible for Enhanced Thermal Stability of Proteins: A Structural Genomics Based Study. *Biochemistry* **41**, 8152-8161.
- Chiong, M., Barra, R., Gonzalez, E., & Vasquez, C. (1988) Resistance of *Thermus* spp. to Potassium Tellurite. *Applied and Environmental Microbiology* **54(2)**, 610612.
- Cicerone, R., & Oremland, R. (1988) Biogeochemical aspects of atmospheric methane. *Global Biogeochemical Cycles* **2**, 299-327.
- Collyn, F., Guy, L., Marceau, M., Simonet, M., & Roten, C.-A. (2006) Describing ancient horizontal gene transfers at the nucleotide and gene levels by comparative pathogenicity island genomics. *Bioinformatics* **22**, 1072-1079.
- Cordeiro, C., Martins, M., Luciano, A., & Silva, R. d. (2002) Production and properties of xylanase from thermophilic *Bacillus* sp. *Brazilian Archives of Biology and Technology* **45**, 413-418.
- Costa, M. D., Rainey, F., & Nobre, M. (2006) The Genus *Thermus* and Relatives. *Prokaryotes* **7**, 797-812.
- Crooks, G., Hon, G., Chandonia, J.-M., & Brenner, S. (2004) WebLogo: A Sequence Logo Generator. *Genome Research* **14**, 1188-1190.
- Darling, A., Miklós, I., & Ragan, M. (2008) Dynamics of genome rearrangement in bacteria populations. *PLOS Genet* **4(7)**, e1000128. doi:10.1371/journal.pgen.1000128.

- Das, S., Paul, S., Bag, S., & Dutta, C. (2006) Analysis of Nanoarchaeum equitans genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. *BMC Genomics* **7**(186), doi:10.1186/1471-2164-7-186.
- Dawson, W., Fukiwara, K., & Kawai, G. (2007). Prediction of RNA pseudoknots using heuristic modeling with mapping and sequencing folding. *PLoS ONE* **2** (9), 3905.
- DeLano, W. (2002, 30th June) The PyMOL molecular graphics system. *The PyMOL molecular graphics system* .
- Dill, K. (1990) Dominant Forces in Protein Folding. *Biochemistry* **29**, 7133-7155.
- Ding, Y., Chan, C., & Lawrence, C. (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Research* **32**, W135-W141.
- Dufraigne, C., Fertil, B., Lespinats, S., Giron, A., & Deschavanne, P. (2005) Detection and characterization of horizontal transfers in prokaryotes using genomic signatures. *Nucleic Acids Research* **33** (1), doi:101093/na/gni004.
- Durme, J., Delgado, J., F., L. S., Schymkowitz, J., Rousseau, F., & Serrano, L. (2011) A graphical interface for the FoldX force field. *Bioinformatics* **27**, 1711-1712.
- Dutta, C., & Pan, A. (2002) Horizontal gene transfer and bacterial diversity. *Journal of Biosciences* **27**, 27-33.
- Edgar, R. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**(5), 1792-1797.
- Erickson, J., & Gross, C. (1989) Identification of sigma E subunit of Escherichia coli RNA polymerase: a second alternate sigma factor involved in high-temperature gene expression. *Gene Dev* **3**, 1462-1471.
- Estell, D., Graycart, T., & Wells, J. (1985) Engineering and enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *The Journal of Biological Chemistry* **260**, 6518-6521.
- Fani, R., & Fondi, M. (2009) Origin and evolution of metabolic pathways. *Physics of Life Reviews* **6**, 23-52.
- Farias, S., & Bonato, M. (2003). Preferred amino acids and thermostability. *Genetics and Molecular Research* , **2**(4), 383-393.
- Farias, S., & Bonato, M. (2002) Preferred codons and amino acid couples in hyperthermophiles. *Genome Biology* **3**(8), 1-18.
- Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. *PHYLIP (Phylogeny Inference Package) version 3.5c* .
- Friedrich, A., Rumszauer, J., Henne, A., & Averhoff, B. (2003) Pillin-Like Protein in the Extremely Thermophilic Bacterium *Thermus thermophilus* HB27: Implication in Competence for Natural Transformation and Links to Type IV Pilus Biogenesis. *Applied and Environmental Microbiology* **69**(7), 3695-3700.

- Froelich, P., Klinkhammer, G., Bender, M., Luedtke, N., Heath, G., Cullen, D., et al. (1979) Early oxidation of organic matter in pelagic sediments of the eastern equatorial Atlantic: suboxic diagenesis. *Geochimica et Cosmochimica Acta* **43**, 1075-1090.
- Ganesan, H., Rakitianskaia, A., Davenport, C., Tummler, B., & Reva, O. (2008) The SeqWord Genome Browser: an online tool for the identification and visualisation of atypical regions of bacterial genomes through oligonucleotide usage. *BMC Bioinformatics* **9**, 333.
- Garg, S., & Johri, B. (1979) Thermophilic moulds in biotechnology. Kluwer Academic Publisher.
- Garrett, S., & Rosenthal, J. (2012) A role for A-to-I RNA editing in temperature adaptation. *Physiology (Bethesda)* **27**, 362-9.
- Giese, M., Betschart, K., Dale, T., Riley, C., Rowan, C., Sprouse, K., et al. (1998) Stability of RNA Hairpins Closed by Wobble Base Pairs. *Biochemistry* **37**, 1094-1100.
- Gogarten, J., Doolittle, W., & Lawrence, J. (2002) Prokaryotic Evolution in Light of Gene Transfer. *Molecular Biology and Evolution* **19(12)**, 2226-2238.
- Goldenfeld, N., & Woese, C. (2007) Biology's next revolution. *Nature* **369**, doi:10.1038/445369a.
- Gómez, B., Fernández-Guerra, A., Casamayor, E., & González, J. M. (2012) Patterns and architecture of genomic islands in marine bacteria. *BMC Genomics* **13**, 347.
- Gounder, K., Brzuszkiewicz, E., Liesegang, H., Wollherr, A., Daniel, R., Gottschalk, G., et al. (2011) Sequence of the hyperplastic genome of the naturally competent *Thermus scotoductus* SA-01. *BMC Genomics* **12**, 577.
- Granick, S. (1957) Speculations on the origins and evolution of photosynthesis. *Annals of the New York Academy of Sciences* **69**, 292-308.
- Groisman, E., Jr, M. S., & Ochman, H. (1992) Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the salmonella genome. *The EMBO Journal* **11(4)**, 1309-1316.
- Gromiha, M., Oobatake, M., & Sarai, A. (1999) Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophysical Chemistry* **82**, 51-67.
- Gromiha, M., Oobatake, M., & Sarai, A. (1999) Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophysical Chemistry* **82**, 51-67.
- Haki, G., & Rakshit, S. (2003) Developments in industrial important thermostable enzymes: a review. *Bioresource Technology* **89**, 17-34.
- Haney, P., Badger, J., Buldak, G., Reich, C., & Woese, C. (1999) Thermal adaptation analysed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *PNAS* **96**, 3578-3583.
- Haney, P., Konisky, J., Koretke, K., Luthey-Schulten, Z., & Wolynes, P. (1997) Structural Basis for Thermostability and Identification of Potential Active Site Residues for Adenylate Kinases From the Archaeal Genus *Methanococcus*. *Proteins: structure, functions and genetics* **28**, 117-130.
- Hayes, W., & Borodovsky, M. (1998) How to interpret an anonymous bacterial genome: Machine learning approach to gene identification. *Genome Research* **8**, 1154-1171.

- Henne, A., Breithaupt, H., Raasch, C., Wiezer, A., Hartsch, T., Liesegang, H., et al. (2004) The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nature Biotechnology* **22**, 547-553.
- Hofacker, I. (2003). Vienna RNA secondary structure server. *Nucleic Acids Research* **31**, 3429-3431.
- Horowitz, N. (1945). On the Evolution of Biochemical Syntheses. *PNAS* **31(6)**, 153-157.
- Hsiao, W., Wan, I., Jones, S., & Brinkman, F. (2003) IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* **19**, 418-420.
- Huang, S.-L., Wu, L.-C., Liang, H.-K., Pan, K.-T., Horng, J.-T., & Ko, M.-T. (2004) PGTdb: a database providing growth temperatures of prokaryotes. *Bioinformatics* **20**, 276-278.
- Hughes, D. (2000) Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. *Genome Biology* **1(6)**, reviews 0006.1–0006.8.
- Jaeger, K.-E., & Reetz, M. (1998) Microbial lipases form versatile tools for biotechnology. *Trends in Biotechnology* **16**, 396-403.
- Jain, R., Rivera, M., & Lake, J. (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Natl. Acad. Sci. USA* **96**, 3801-3806.
- Jensen, R. (1976) Enzyme Recruitment in Evolution of New Function. *Annu. Rev. Microbiology* **30**, 409-425.
- Jónsson, Z., Thorbjarnardóttir, S., & Palsdóttir, A. (1994) Sequence of the DNA ligase-encoding gene from *Thermus scotoductus* and conserved motifs in DNA ligases. *Gene* **151**, 177-180.
- Kanehisa, M., & Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28(1)**, 27-30.
- Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K., Itoh, M., Kawashima, S., et al. (2006) From Genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* **34**, Database issue doi:10.1093/nar/gkj102
- Kannan, N., & Visheeshwara, S. (2000) Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Engineering* **13**, 753-761.
- Karlin, S., & Burge, C. (1995) Dinucleotide relative abundance extremes: a genomic signature. *TIG* **11(7)**, 283-290.
- Karlin, S., Campbell, A., & Mrázek, J. (1998a) Comparative DNA analysis across diverse genomes. *Annu. Rev. Genet.* **32**, 185-225.
- Karlin, S., Mrazek, J., & Campbell, A. (1998) Codon usage in different gene classes of the *Escherichia coli* genome. *Molecular Microbiology* **29(6)**, 1341-1351.
- Karlin, S., Mrázek, J., & Campbell, A. (1996) Frequent oligonucleotides and peptides of the *Haemophilus influenzae* genome. *Nucleic Acids Research* **24(21)**, 4263-4272.
- Karp, P., Paley, S., & Romero, P. (2002) The Pathway tools software. *Bioinformatics* **18(1)**, S1-S8.
- Katoh, K., Misawa, K., & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066.

- Kieft, T., Fredrickson, J., Onstott, T., Gorby, Y., Kostandarithes, H., Bailey, T., et al. (1999) Dissimilatory Reduction of Fe(III) and Other Electron Acceptors by a *Thermus* Isolate. *Applied and Environmental Microbiology* **65**(3), 1214-1221.
- Knudsen, B., & Hein, J. (2003) Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Research* **31**, 3423-3428.
- Kolesov, G., Wunderlich, Z., Laikova, O., Gelfand, M., & Mirny, L. (2007) How gene order is influenced by biophysics of transcription regulation. *PNAS* **104**, 13948-13953.
- Kristjansson, J. (1989) Thermophilic organisms as source of thermostable enzymes. *Trends in Biotechnology* **7**, 349-353.
- Kumar, S., Ma, B., Tsai, C.-J., Sinha, N., & Nussinov, R. (2000) Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Science* **9**, 10-19.
- Kumar, S., Tsai, C., & Nussinov, R. (2000) Factors enhancing protein thermostability. *Protein Engineering* **13**(3), 179-191.
- Kurosowa, N., Itoh, Y., & Itoh, T. (2005) *Thermus Kawarayensis* sp. nov., a new member of the genus *Thermus*, isolated from Japanese hot springs. *Extremophiles* **9**, 81-84.
- Langille, M., & L., F. S. (2009) IslandViewer: an integrated interface for computational identification and visualisation of genomic islands. *Bioinformatics* **25**(5), 664-665.
- Lasa, I., & Berenguer, J. (1993) Thermophilic enzymes and their biotechnology potential. *Microbiologia* **9**, 77-89.
- Lawrence, G., & Ochman, H. (1997) Amelioration of Bacterial Genomes: Rates of Change and Exchange. *Journal of Molecular Evolution* **44**, 383-397.
- Lazcano, A., & Miller, S. (1996) The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell* **85**, 793-798.
- Lemaitre, C., Tannier, E., Gautier, C., & Sagot, M.-F. (2008) Precise detection of rearrangement breakpoints in mammalian chromosomes. *BMC Bioinformatics* **9**, 286.
- Leuschner, C., & Antranikian, G. (1995) Heat-stable enzymes from extremely thermophilic and hyperthermophilic microorganisms. *World Journal of Microbiology & Biotechnology* **11**, 91-114.
- Lin, Y., Lu, C., Liu, Y.-C., & Tang, C. (2006) SPRING: a tool for the analysis of genome rearrangement using reversals and block-interchange. *Nucleic Acids Research* **34**, Web Server issue doi:10.1093/nar/gkl169
- Lioliou, E., Pantazaki, A., & Kyriakidis, D. (2004) *Thermus thermophilus* genome analysis: benefits and implications. *Microbial Cell Factories* **3**(5), 1723-1727.
- López de Paz, M., Lacroix, E., Ramirez-Alvarado, M., & Serrano, L. (2001) Computer-aided design of beta sheet peptides. *J. Mol. Biol.* **312**, 229-246.
- Loveley, D., Baedeker, M., Lonergan, D., Cozzarelli, I., Phillips, E., & Siegel, D. (1989) Oxidation of aromatic contaminants coupled to microbial iron reduction. *Nature* **339**, 297-300.
- Lovley, D. (1991) Dissimilatory Fe(III) and Mn(IV) Reduction. *Microbiological Reviews* **55**(2), 259-287.

- Madden, T. (2002, October) The BLAST Sequence Analysis Tool. *The BLAST Sequence Analysis Tool* .
- Majeti, N., & Kumar, R. (2000) A review of chitin and chitosan applications. *Reactive Functional Polymers* **46**, 1-27.
- Markham, N., & Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. In *Bioinformatics* (Ed.). Totowa, NJ.: Human Press.
- Marri, P., & Golding, G. (2008) Gene amelioration demonstrated: the journey of nascent genes in bacteria. *Genome* **51**, 164-168.
- Martin-Renon, M., Stuart, A., Fisher, A., Sanchez, R., Melo, F., & Sali, A. (2000) Comparative protein structure modelling of genes and genomes. *Anu. Rev. Biophys. Biomol.* **29**, 291-325.
- Matsui, I., Sakai, Y., Matsui, E., Kukuchi, H., Kawarabayasi, Y., & Honda, K. (2000) Novel substrate specificity of a membrane-bound β -glycosidase from the hyperthermophilic archaeon *Pyrococcus horikoshii*. *FEBS Letters* **467**, 195-200.
- Mattos, C. (2002) Protein-water interactions in dynamic world. *Trends in Biochemical Sciences* **27(4)**, 203-208.
- Mazurie, A., Bonchev, D., Schwikowski, B., & Buck, A. (2010) Evolution of metabolic network organisation. *BMC Systems Biology* **4(59)**, 1752-1759.
- McDonald, J. (2010) Temperature adaptation at homologous sites in proteins from nine thermophile-mesophile species pairs. *Genome Biol. Evol.* **2**, 267-76.
- McDonald, J., Grasso, A., & Rejto, L. (1999) Patterns of temperature adaptation in proteins from *Methanococcus* and *Bacillus*. *Mol. Biol. Evol.* **16**, 1785-1970.
- Miguel, T. d., Barros-Velazquez, J., & Vila, T. (2006) Industrial Applications of Hyperthermophilic Enzymes: A Review. *Protein & Peptide Letters* **13**, 645-651.
- Mohsen, A., Khader, A., Ramachandram, D., & Ghallab, A. (2010) Predicting the minimum free energy RNA Secondary Structures using Harmony Search Algorithm. *International Journal of Biological and Life Sciences* **6(3)**, 157-163.
- Mullis, K., Faloona, F., Scharf, S., Saiki, T., Horn, G., & Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **L1**, 263-273.
- Mumford, E. (1913) New Iron Bactekm. *J. Chem. Soc.* **103**, 645-650.
- Munster, M., A.P., M., Woodrow, J., & Sharp, R. (1986) Isolation and Preliminary Taxonomic Studies of *Thermus* Strains Isolated from Yellowstone National Park, USA. *International Journal of Systematics Bacteriology* **132**, 1677-1683.
- Nadeau, J., & Taylor, B. (1984) Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc. Natl. Acad. Sci. USA* **81**, 814-818.
- Nagi, A., Anderson, K., & Regan, L. (1999) Using loop length variants to dissect the folding pathways of a four-helix-bundle protein. *J. Mol. Biol.* **286**, 257-265.

- Nicolas, P., Bize, L., Muri, F., Hoebeker, M., Rodolphe, F., Ehrlich, S., et al. (2002) Mining *Bacillus subtilis* chromosome heterogeneities using hidden Markov models. *Nucleic Acids Research* **30**(6), 1418-1426.
- Niehaus, F., Bertoldo, C., Kahler, M., & Antranikian, G. (1999). Extremophiles as a source of novel enzymes for industrial application. *Applied Microbiol Biotechnology* **51**, 711-729.
- Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdahary, B., & Landegren, U. (1994). Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085-2088.
- Nishio, Y., Nakamura, Y., Kawarabayasi, Y., Usuda, Y., Kimura, E., & Sugimoto, S. (2003) Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. *Genome Research* **13**, 1572-1579.
- Ogata, H., Wataru, F., Goto, S., & Kanehisa, M. (2000) A heuristics graph comparison algorithm and its application to detect functionally related enzyme clusters. *Nucleic Acids Research* **28**, 4021-4028.
- Opperman, D., & Heerden, E. v. (2007) Aerobic Cr(VI) reduction by *Thermus scotoductus* strain SA-01. *Journal of Applied Microbiology* **103**, 1907-1913.
- Oshima, T., & Imahori, K. (1974) Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a Nonsporulating Thermophilic Bacterium from a Japanese Thermal Spa. *International Journal of Systematics Bacteriology* **24**(1), 102-112.
- Overbeek, R., Larsen, N., Pusch, G., D'Souza, M., Jr, E. S., Kyrpides, N., et al. (2000) WIT: integrated system for high throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Research* **28**, 123-125.
- Pack, S., & Yoo, Y. (2005) Packing-based differences of structural features between thermophilic and mesophilic proteins. *International Journal of Biological Macromolecules* **35**, 169-174.
- Pack, S., & Yoo, Y. (2004) Protein thermostability: structure-based difference of amino acid between thermophilic and mesophilic. *Journal of Bacteriology* **111**, 269-277.
- Palmenaer, D., Siguer, P., & Mahillon, J. (2008) IS4 family goes genomic. *BMC Evolutionary Biology* **8**, DOI:10.1186/1471-2148-8-18.
- Panasik, N., Brenchley, J., & Farber, G. (2000) Distribution of structural features contributing to thermostability in mesophilic and thermophilic alpha/beta barrel glycosyl hydrolases. *Biochim Biophys Acta* **1543**, 189-201.
- Parisen, M., & Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51-55.
- Pask-Hughes, R., & Williams, R. (1975) Extremely Thermophilic Gram-negative Bacteria from Hot Tap Water. *Journal of General Microbiology* **88**, 321-328.
- Paul, J. (1999) Microbial Gene Transfer: An Ecological Perspective. *J. Molec. Microbiol. Biotechnol.* **1**, 45-50.
- Pevzner, P., & Tesler, G. (2003) Genome Rearrangements in Mammalian Evolution: Lessons From Human and Mouse Genomes. *Genome Research* **13**, 37-45.

- Pie, J., Kim, B.-H., & Grishin, N. (2008) PROMALS3D: a tool for multiple sequence and structure alignment. *Nucleic Acids Research* **36**, 2295-2300.
- Ponnuswamy, P., Muthusamy, R., & Manavalan, P. (1982) Amino acid composition and thermal stability of proteins. *International Journal of Biological Macromolecules* **4(3)**, 186-190.
- Rao, M., Tanksale, A., Ghatge, M., & Deshpande, V. (1998) molecular and biotechnology aspects of microbial proteases. *Microbiology and Molecular Biology Reviews* **62**, 597-635.
- Reva, O., & Tummler, B. (2005) Differentiation of regions with atypical oligonucleotide composition in bacterial genomes. *BMC Bioinformatics* **6**, 251.
- Reva, O., & Tummler, B. (2004) Global features of sequences of bacterial chromosomes, plasmids and phages revealed by analysis of oligonucleotide usage patterns. *BMC Bioinformatics* **5**, 90.
- Rieseberg, L. (2001) Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution* **16**, 351-358.
- Rison, S., & Thornton, J. (2002) Pathway evolution, structurally speaking. *Current Opinion in Structural Biology* **12**, 374-382.
- Rocha, E., Guerdoux-Jamet, P., Moszer, I., Viari, A., & Danchin, A. (2000) Implications of gene distribution in the bacterial chromosome for the bacteria cell factory. *Journal of Biotechnology* **78**, 209-219.
- Rowbotham, A., Levy, S., & Shuker, L. (2010) Chromium in the environment: An evaluation of exposure of the UK general population and possible adverse health effects. *Journal of Toxicology and Environmental Health* **3(3)**, 145-178.
- Roycroft, A., Thompsom, G., & Hammond, S. (2006) The role of cell surface polysaccharide antigens in the pathogenicity of Escherichia coli. *FEMS Microbiology Letters* **18**, 49-53.
- Russell, R., Ferguson, J., Hough, D., Danson, M., & Taylor, G. (1997) The Crystal Structure of Citrate Synthase from the Hyperthermophilic Archaeon Pyrococcus furiosus at 1.9 Amstrong Resolution. *Biochemistry* **36**, 9983-9994.
- Russell, R., Gerike, U., Danson, M., Hough, D., & Taylor, G. (1998) Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure* **6**, 351-361.
- Sadeghi, M., Naderi-Manesh, H., Zarrabi, M., & Ranjbar, B. (2006) Effective factors in thermostability of thermophilic proteins. *Biophysical Chemistry* **119**, 256-270.
- Saiki, T., Kimura, R., & Arima, K. (1972). Isolation and Characterisation of Extremely Thermophilic Bacteria from Hot Springs. *Agri. Biol. Chem.* **36 (13)**, 2366-1972.
- Sali, A., & Blundell, T. (1993) Comparative protein modelling by satisfaction of spartial restraints. *J. Mol. Biol.* **234**, 779-815.
- Sanchez, I. (2008) Protein folding transtion states probed by loop extension. *Protein Science* **17**, 183-186.
- Scandurra, R., Consalvi, V., Chiaraluze, R., Politi, L., & Engel, P. (1998) Protein thermostability in extremophiles. *Biochimie* **80**, 933-941.

- Schalling, M., Hudson, T., Buetow, K., & Housman, D. (1993) Direct detection of novel expanded trinucleotide repeats in the human genome. *Nature Genetics* **4**, 135-139.
- Schwarzenlander, C., Haase, W., & Averhoff, B. (2009) The role of single subunits of the DNA transport machinery of *Thermus thermophilus* HB27 in DNA binding and transport. *Environmental Microbiology* **11**(4), 801-808.
- Schymkowitz, J., R., F. N., Rousseau, F., & Serrano, L. (2005) The FoldX webserver: an online force field. *Nucleic Acids Research* **33**, W382-8.
- Seeliger, D., & Groot, B. d. (2010) Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophysical Journal* **98**, 2309-2316.
- Shahidi, F., Arachchi, J., & Jeon, Y.-J. (1999) Food applications of chitin and chitosans. *Trends in Food Science & Technology* **10**, 37-51.
- Shapiro, S., & Wilk, M. (1965) An analysis of variance test for normality. *Biometrika* **52**, 591.
- Sharp, P., & Li, W.-H. (1987). The codon Adaptation Index -a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* **15**, 1281-1295.
- Shen, M., & Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Science* **15**, 2507-2524.
- Simeonidis, E., Rison, S., Thornton, J., Bogle, I., & Parpageorgiou, L. (2003) Analysis of metabolic networks using a pathway distance metric through linear programming. *Metabolic Engineering* **5**, 211-219.
- Skirnisdottir, S., Hreggvidsson, G., Holst, O., & Kristjansson, J. (2001) Isolation and characterization of a mixotrophic sulfur-oxidizing *Thermus scotoductus*. *Extremophiles* **5**, 45-51.
- Snel, B., Bork, P., & Huynen, M. (1999) Genome Phylogeny based on gene content. *Nature Genetics* **21**, 108-110.
- Soding, J., Biegert, A., & Lupas, A. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* **33**, doi:10.1093/nar/gki408.
- Spirin, V., Gelfand, M., Mironov, A., & Mirny, L. (2006) A metabolic network in the evolutionary context: Multiscale structure and modularity. *PNAS* **103**(23), 8774-8779.
- Stankiewicz, P., & Lupski, J. (2002) Genome architecture, rearrangements and genomic disorders. *Trends in Genetics* **18**, 74-82.
- Synowiecki, J. (2010) Some applications of thermophiles and their enzymes for protein processing. *African Journal of Biotechnology* **9**, 7020-7025.
- Szilagyi, A., & Zavodszky, P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**, 493-504.
- Takayama, K., & Kjelleberg, S. (2000) The role of RNA stability during bacteria stress responses and starvation. *Environmental Microbiology* **2**, 355-365.
- Talavera, G., & Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systems Biology* **56**, 564-577.

- Tomazic, S., & Klibanov, A. (1988) Mechanisms of Irreversible Thermal Inactivation of Bacillus alpha -amylases. *The Journal of Biological Chemistry* **263** (7), 3086-3091.
- Tophan, C., Srinivasan, N., & Blundell, T. (1997) Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Engineering* **10**, 7-21.
- Trivedi, S., Gehlot, H., & Rao, S. (2006). Protein thermostability in Archea and Bacteria. *Genetics and Molecular Research* **5** (4), 816-827.
- Tsai, C.-J., & Nussinov, R. (1997) Hydrophobic folding units at protein-protein interfaces: implications to protein folding to protein-protein association. *Protein Science* **6**, 1426-1437.
- van den Burg, B., Vriend, G., Veltman, O., Venema, G., & Eijsink, V. (1998) Engineering an enzyme to resist boiling. *PNAS* **96**, 2066-2060.
- van der Linden, M., & Farias, S. (2006) Correlation between codon usage and thermostability. *Extremophiles* **10**, 479-481.
- van der Linden, M., Rego, T., Aaujo, D., & Farias, S. (2006) Prediction of potential thermostable proteins in *Xylella fastidiosa*. *Journal of Theoretical Biology* **242**, 421-425.
- Veltman, O.R., Vriend, G., Middlehoven, P.J., van den Burg, V., Venema, G., & Eijsink, V.G.H. (1996) Analysis of structural determinants of the stability of thermolysin-like proteases by molecular modelling and site-directed mutagenesis. *Protein Engineering* **9**, 1181-1189.
- Vogt, G., Woell, S., & Argos, P. (1997) Protein Thermal Stability, Hydrogen Bonds, and Ion Pairs. *Journal of Molecular Biology* **269**, 631-643.
- Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W., et al. (2006) Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* **7** (142), doi:10.1186/1471-2105-7-142.
- Wang, B. (2001) Limitations of compositional approach to identifying horizontally transferred genes. *Journal of Molecular Evolution* **53**, 244-250.
- Watanabe, K., Hata, Y., Kizaki, H., Katsube, Y., & Suzuki, Y. (1997) The Refined Crystal Structure of Bacillus cereus oligo-1-6-glucosidase at 2.0 Å Resolution: Structural Characterisation of Proline-substitution Sites for Protein Thermostabilization. *Journal of Molecular Biology* **269**, 142-153.
- Watterson, G., Ewens, W., & Hall, T. (1982) The Chromosome Inversion Problem. *Journal of Theoretical Biology* **99**, 1-7.
- Williams, R., Smith, K., Welch, S., Micallef, J., & Sharp, R. (1995) DNA Relatedness of Thermus Strains, Description of Thermus brockianus sp. nov., and Proposal to Reestablish Thermophilus (Oshima and Imahori). *International Journal of Systematics Bacteriology* **45** (3), 495-499.
- Wolf, Y., Rogozin, I., & Kondrashov, A. (2001) Genome Alignment, Evolution of Prokaryotic Genome Organization, and Prediction of Gene Function Using Genomic Context. *Genome Research* **11**, 356-372.
- Xia, T., Jr, J. S., Burkard, M., Kierzek, R., Schoeder, S., Jiao, X., et al. (1998) Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs. *Biochemistry* **37**, 14719-14735.

- Xu, Z., Liu, Y., Yang, Y., Jiang, W., Arnold, E., & Ding, J. (2003) Angstroms: Insights into the Molecular Burkholderia pickettii at a Resolution of 2.7 Crystal Structure of d-Hydantoinase from Basis of Enzyme Thermostability. *Journal of Bacteriology* **185**, 4038-4029.
- Ycas, M. (1974) On Earlier States of the Biochemical Systems. *Journal of Theoretical Biology* **44**, 145-160.
- Yin, Y., Zhang, H., Olman, V., & Xu, Y. (2010) Genomic arrangement of bacterial Operons is constrained by biological pathways encoded in the genome. *PNAS* **107** (14), 6310-6315.
- Yip, K., Stillman, T., Britton, K., Artymiuk, P., Baker, P., Sedelnikova, S., et al. (1995) The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**, 1147-1158.
- Zhou, X.-X., Wang, Y.-B., Pan, Y.-J., & Li, W.-F. (2008) Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. *Amino Acids* **34**, 25-33.
- Zuker, M., & Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxilliary information. *Nucleic Acids Research* **9** (1), 133-147.

Appendices

6.1 Appendix A

6.1.1 Pathway and Breakpoints Results

Table 6.1.1: Number of breakpoints affecting pathways

Genome	Pathways with Breakpoints	Number of Breakpoints	Total Breakpoints	Total Insertions
SA-01	Nitrate reduction (dissimilatory)	1	290	1486
	purine nucleotides de novo biosynthesis II	1		
	super pathway of leucine, valine, and isoleucine biosy.	1		
	superpathway of fatty acid biosynthesis I	1		
HB8	purine nucleotides de novo biosynthesis II	2	505	1260
	threonine biosynthesis	1		
	glycolysis II	1		
	superpathway of fatty acid biosynthesis I	1		
	arginine biosynthesis II (acetyl cycle)	1		
tetrapyrrole biosynthesis I	4			
HB27	purine nucleotides de novo biosynthesis II	2	495	1237
	threonine biosynthesis	1		
	glycolysis II	1		
	superpathway of fatty acid biosynthesis I (<i>E. coli</i>)	1		
	arginine biosynthesis II (acetyl cycle)	1		
tetrapyrrole biosynthesis I	4			

6.2 Appendix B

6.2.1 Common Pathway Clustering

Table 6.2.1: Clustering of thirty-eight common pathways

Common Pathways	Number of Genes						
	SA-01	HB 8	HB27	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. ruber</i>	<i>M. silvanus</i>
superpathway of leucine, valine, and isoleucine biosynthesis	9	13	15	14	13	12	13
arginine biosynthesis I	12	10	11	12	13	12	13
histidine biosynthesis	9	6	8	8	8	6	8
isoleucine biosynthesis I (from threonine)	5	5	8	10	5	5	5
leucine biosynthesis	9	13	15	14	13	12	13
ornithine biosynthesis	6	5	6	6	8	12	8
tryptophan biosynthesis	6	6	8	19	7	7	7
valine biosynthesis	5	4	5	8	4	4	4
tRNA charging pathway	71	67	73	110	71	71	71
chorismate biosynthesis I	9	6	8	11	7	8	7
gluconeogenesis I	11	6	14	18	11	12	11
UDP-N-acetylmuramoyl-pentapeptide biosynthesis III	8	7	7	10	5	4	5
peptidoglycan biosynthesis III	10	9	9	12	7	6	7
O-antigen biosynthesis (<i>E. coli</i>)	6	4	4	10	11	11	11
pantothenate and coenzyme A biosynthesis I	7	4	5	8	11	11	11
methylerythritol phosphate pathway	8	6	8	8	8	8	8
adenosylcobalamin biosynthesis II (late cobalt incorporation) biosynthesis	9	7	13	6	5	7	5
formylTHF biosynthesis I	6	6	8	11	7	6	7
tetrahydrofolate biosynthesis	9	6	8	15	8	5	8
folate transformations	8	7	7	5	7	6	7
5-aminoimidazole ribonucleotide biosynthesis I	7	6	6	5	7	7	7
inosine-5'-phosphate	85	4	5	5	5	5	5

biosynthesis II							
guanosine nucleotides de novo biosynthesis	10	5	8	8	4	7	4
adenosine nucleotides de novo biosynthesis	9	7	7	8	5	7	5
purine nucleotides de novo biosynthesis II	16	12	15	13	15	15	15
salvage pathways of purine and pyrimidine nucleotides	17	14	17	16	15	14	15
salvage pathways of adenine, hypoxanthine, and their nucleosides	5	5	5	10	4	4	4
uridine-5'-phosphate biosynthesis	6	5	6	7	5	5	5
pyrimidine ribonucleotides de novo biosynthesis	9	8	10	11	8	7	8
salvage pathways of pyrimidine ribonucleotides	7	5	5	7	7	6	7
pyrimidine deoxyribonucleotides de novo biosy.	9	6	7	9	4	7	4
fatty acid β-oxidation I	13	14	15	9	3	3	3
homolactic fermentation	13	10	16	22	12	13	12
mixed acid fermentation	10	5	9	26	9	10	9
glycolysis III (Thermotoga)	9	7	12	17	10	10	10
glycolysis I	9	7	12	17	10	10	10
glycolysis II	9	7	12	17	10	10	10
respiration (anaerobic)	9	5	9	18	9	11	9

6.3 Appendix C

6.3.1 Horizontal gene transfer stratigraphic analysis

Table 6.3.1: Tree topologies for orthologous genes

Recipient	Donor	Gene ID	Left border	Right border	Direction	Gene name	Product or function	Consensus tree for the group
Group 1								
SA-01	M.ruber	TSC_c00670	52807	53553	Rev	argB1	acetylglutamate kinase	
M.ruber	unknown	TSC_c01140	97267	98151	Rev		Duf1395 protein	
M.ruber	unknown	TSC_c01350	119514	120131	Dir	rnhB	ribonuclease CBS domain containing protein	
SA-01	unknown	TSC_c01460	130751	131182	Dir			
SA-01	M.ruber	TSC_c01530	136508	137917	Rev	gatB	glutamyl-tRNA(Gln) translation elongation factor Tu	
M.ruber	unknown	TSC_c01740	150776	151996	Dir	tuf1		
SA-01	M.ruber	TSC_c01780	152982	153425	Dir	rplK	ribosomal protein 50S	
SA-01	M.ruber	TSC_c01800	154312	154833	Dir	rplJ	ribosomal protein 50S	
SA-01	M.ruber	TSC_c01810	154844	155221	Dir	rplL	ribosomal protein 50S	
SA-01	M.ruber	TSC_c01820	155312	155893	Dir		phenylacrylic acid decarboxylase	
M.silvanus	unknown	TSC_c01940	163915	164724	Dir	argB2	acetylglutamate kinase	
HB27	unknown	TSC_c02220	187412	187717	Dir	rpsF	ribosomal protein 30S	
M.silvanus	unknown	TSC_c02240	188536	188802	Dir	rpsR	ribosomal protein 30S	

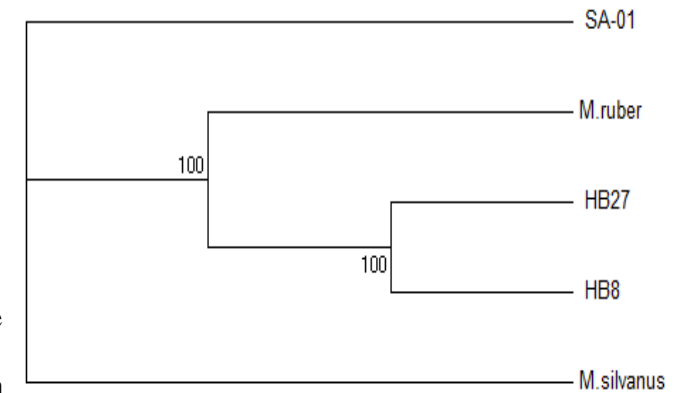
SA-01	M.ruber	TSC_c03570	306322	306699	Dir		S1 domain protein
M.silvanus	unknown	TSC_c06250	572669	572971	Dir		nitrogen regulatory protein P-II
M.ruber	unknown	TSC_c07450	687166	689112	Dir	acs2	acetyl-CoA synthetase
SA-01	M.ruber	TSC_c12700	1205913	1207904	Dir	tmk1	thymidylate kinase
SA-01	unknown	TSC_c12710	1208026	1209228	Dir		beta-ketoadipyl-CoA thiolase
HB8	unknown	TSC_c13850	1332430	1332891	Rev	rplS	ribosomal protein 50S
SA-01	M.ruber	TSC_c15000	1442307	1443050	Rev	aapP	general L-amino acid transport, ATP-binding conserved hypothetical protein
SA-01	M.ruber	TSC_c15120	1456131	1456892	Rev		excinuclease ABC subunit C
SA-01	M.ruber	TSC_c18650	1767588	1769366	Dir	uvrC	cell division protein
M.ruber	unknown	TSC_c20260	1919211	1921094	Rev	ftsH2	conserved hypothetical protein
SA-01	M.ruber	TSC_c20380	1931310	1931729	Dir		response regulator
SA-01	unknown	TSC_c21620	2044185	2044859	Dir		ribosomal protein 50S
SA-01	M.ruber	TSC_c22370	2101643	2102068	Rev	rplP	transcriptional regulator, GntR family
SA-01	unknown	TSC_c22680	2126247	2126909	Rev		cehydrogenase
SA-01	unknown	TSC_c22830	2144164	2144847	Dir		methylmalonyl-CoA epimerase
SA-01	M.ruber	TSC_c22840	2145147	2145548	Rev	mce	polyferredoxin
SA-01	M.ruber	TSC_c22860	2147353	2148294	Rev		phospho-2-dehydro-3-deoxyheptonate aldolase
SA-01	M.ruber	TSC_c22870	2148295	2149359	Rev	aroF2	twin arginine-targeting protein translocase
SA-01	unknown	TSC_c22880	2149393	2150136	Rev	tatC	maltose ABC transporter, permease protein
SA-01	unknown	TSC_c23380	2197169	2198497	Dir		maltodextrin glucosidase
M.silvanus	unknown	TSC_c23400	2199822	2201534	Dir		

s

M.ruber	unknown	TSC_c23420	2201843	2203654	Dir	mutS2	DNA mismatch repair protein MutS
SA-01	M.ruber	TSC_c24100	2262279	2262962	Rev		conserved hypothetical protein
SA-01	M.ruber	TSC_c24930	2335644	2336099	Dir		conserved hypothetical protein

Group 2

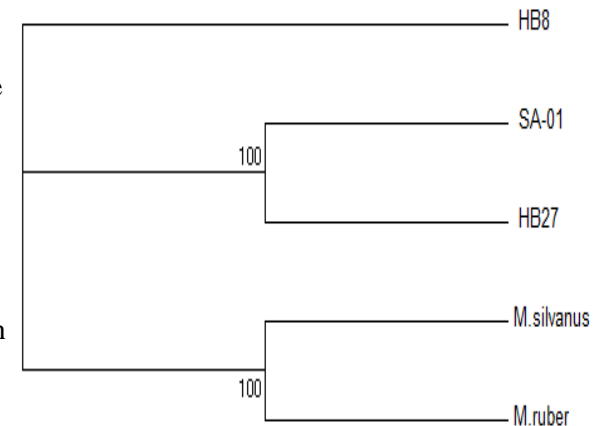
SA-01	M.silvanus	TSC_c01550	139048	141714	Rev	pilF	type IV pilus assembly protein conserved hypothetical protein
SA-01	M.silvanus	TSC_c01580	142228	142953	Rev		transporter 3-hydroxyisobutyrate dehydrogenase
SA-01	M.silvanus	TSC_c01600	143912	144295	Dir	crcB	
SA-01	M.silvanus	TSC_c02150	181419	182561	Rev		mechanosensitive ion channel
SA-01	M.silvanus	TSC_c02310	192485	193354	Dir		
SA-01	M.silvanus	TSC_c02500	210696	211751	Rev		chloromuconate cycloisomerase
M.silvanus	unknown	TSC_c04090	352392	353471	Dir		
M.silvanus	SA-01	TSC_c04530	398877	399551	Dir		two-component response regulator
M.ruber	unknown	TSC_c05050	456122	456835	Rev		branched-chain amino acid ABC transporter, ATP-binding protein
M.silvanus	unknown	TSC_c05650	518938	519666	Dir		
SA-01	M.silvanus	TSC_c05980	548855	549973	Rev	nuoH	NADH-quinone
SA-01	M.silvanus	TSC_c12840	1223727	1225076	Rev		aminotransferase



SA-01 M.silvanus	M.ruber	TSC_c14570	1400143	1400643	Dir	ruvC	crossover junction endodeoxyribonuclease
SA-01	unknown	TSC_c14860	1428806	1429699	Dir		transporter
SA-01	M.silvanus	TSC_c15590	1499533	1502664	Dir	ileS	isoleucine-tRNA tRNA pseudouridine synthase D
SA-01	M.silvanus	TSC_c18590	1763664	1764719	Dir	truD	iron-sulfur cluster- binding protein
HB8 M.silvanus	unknown	TSC_c19550	1854592	1855740	Rev		conserved hypothetical protein
SA-01	unknown	TSC_c19860	1880814	1881308	Rev		
SA-01	unknown	TSC_c22710	2129094	2130281	Dir	bioF	8-amino-7-oxononanoate zinc-binding dehydrogenase
SA-01	M.silvanus	TSC_c22770	2136825	2137859	Dir		
SA-01	unknown	TSC_c22800	2139609	2140652	Dir		hypothetical membrane spanning protein methionine adenosyltransferase
SA-01	M.silvanus	TSC_c23310	2191412	2192590	Rev	metK	
SA-01	M.silvanus	TSC_c23750	2232596	2233294	Dir		carboxylesterase

Group 3

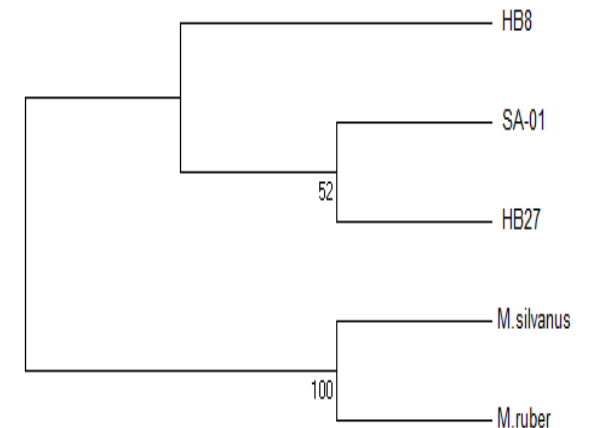
HB8	unknown	TSC_c01230	106764	107159	Dir		ferric uptake regulatory protein
HB8	unknown	TSC_c01240	107182	108423	Dir		folyl-polyglutamate synthetase
HB8	unknown	TSC_c01250	108451	109581	Rev		aminotransferase molybdopterin biosynthesis mog protein
HB8	unknown	TSC_c01260	109574	110068	Rev		conserved hypothetical protein
HB8	unknown	TSC_c07600	700408	700941	Rev		
HB8	unknown	TSC_c07680	705384	706397	Dir	trpS	tryptophane-tRNA



HB8	unknown	TSC_c08750	822804	823403	Dir		ATP synthase, subunit (E/31 kDa)
HB8	unknown	TSC_c09640	918492	920636	Rev	pnp	polyribonucleotide nucleotidyltransferase shikimate/quininate 5-dehydrogenase
SA-01	unknown	TSC_c09760	929656	930558	Dir		5-carboxymethyl-2-hydroxymuconate
M.ruber	unknown	TSC_c10440	987993	989543	Dir	hpaE	glutamyl-tRNA(Gln)
HB27	M.ruber	TSC_c13320	1282190	1282468	Dir	gatC	seryl-tRNA
HB27	M.ruber	TSC_c13330	1282471	1283739	Dir	serS	mannitol-binding protein metallo-beta-lactamase family protein
HB8	unknown	TSC_c15490	1488849	1489931	Dir		L-ribulose-phosphate 4-epimerase
HB8	unknown	TSC_c17180	1628099	1629535	Dir		acetoin utilization protein
HB8	unknown	TSC_c17780	1683922	1684983	Dir		endonuclease V
HB8	unknown	TSC_c19660	1861749	1862873	Rev	acuC	conserved hypothetical protein
HB8	M.ruber	TSC_c20230	1917321	1917998	Dir	nfi	elongation factor G
HB8	unknown	TSC_c20580	1953393	1953806	Dir		
HB8	M.ruber	TSC_c21580	2038176	2040176	Dir		

Group 4

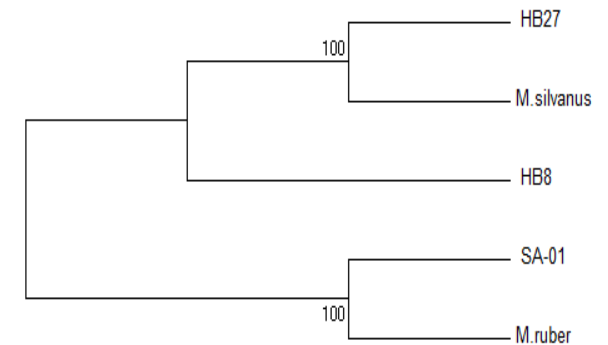
HB27	unknown	TSC_c01920	161964	162860	Dir	lysX	lysine biosynthesis enzyme
M.ruber	HB27	TSC_c03280	283033	283575	Rev		probable acetyltransferase dNTP
SA-01	HB8	TSC_c05820	535187	536317	Rev		triphosphohydrolase
HB27	unknown	TSC_c09270	880628	881503	Rev	czrB	cadmium-zinc resistance protein
HB8	M.silvanus	TSC_c09880	938608	939546	Rev		ferric enterobactin esterase related protein



M.ruber	HB27	TSC_c10220	968590	969711	Dir		ABC transporter periplasmic binding protein
HB27	unknown	TSC_c13470	1297402	1298880	Rev	hemL	glutamate-1- semialdehyde-2,1- aminomutase
M.silvanus	unknown	TSC_c13480	1299292	1300287	Rev	pdhB	pyruvate pyruvate dehydrogenase E1 beta- subunit
M.ruber	HB27	TSC_c13610	1311238	1312407	Rev	pgk	phosphoglycerate kinase imidazoleglycerol phosphate synthase, cyclase subunit
M.silvanus	unknown	TSC_c18720	1776306	1777085	Dir	hisF	lipoprotein releasing system transmembrane protein
M.silvanus	HB27	TSC_c22510	2110733	2111854	Rev		
M.ruber	unknown	TSC_c22540	2112640	2114205	Rev		Dak phosphatase glutamine-fructose-6- phosphate transaminase
HB27	unknown	TSC_c24230	2276748	2278562	Dir	glmS2	

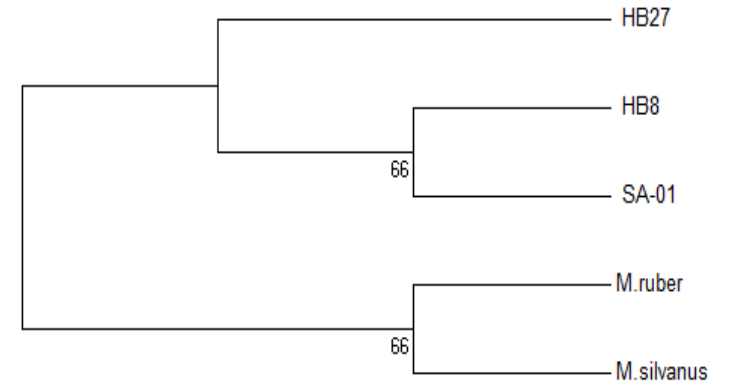
Group 5

HB8	unknown	TSC_c01790	153418	154107	Dir	rplA	ribosomal protein 50S carbon-nitrogen hydrolase family protein
SA-01	M.ruber	TSC_c01960	166317	167183	Dir		CoA-binding protein
HB27	M.silvanus	TSC_c01980	168022	168435	Dir		



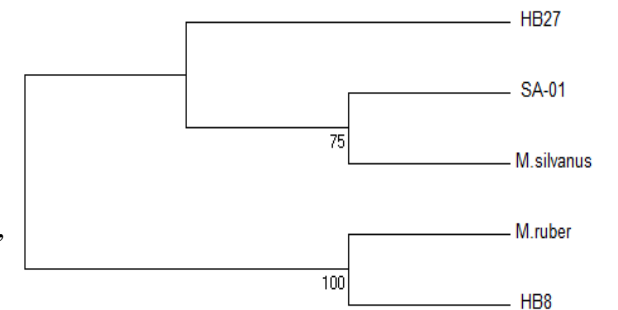
Group 6

HB27	M.ruber	TSC_c00820	68149	68454	Rev	groS	chaperonin transporter
HB8	unknown	TSC_c04050	349029	349865	Rev		
HB27	unknown	TSC_c05080	458933	459901	Rev	hmgE	



Group 7

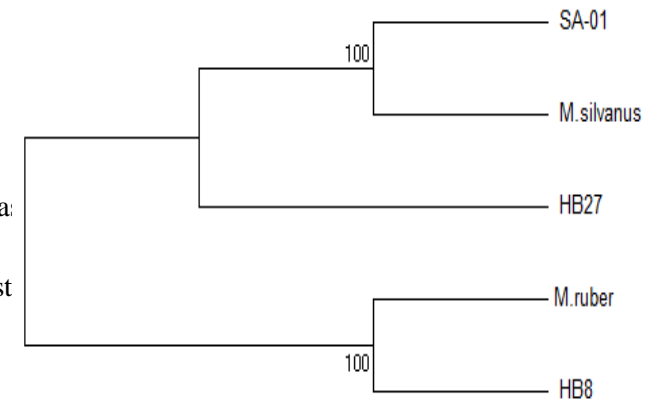
HB8	unknown	TSC_c01280	111143	111733	Dir		acetyltransferase, gnat family 4-hydroxyphenylacetate degradation bifunctional isomerase/decarboxylase, C-subunit
HB8	M.ruber	TSC_c10430	987256	987996	Dir	hpaG	



Group 8

HB27	unknown	TSC_c23290	2188357	2190249	Dir	speA	arginine decarboxylase, diguanylate cyclase/phosphodiesterase
HB8	M.ruber	TSC_c23300	2190296	2191372	Dir		

147

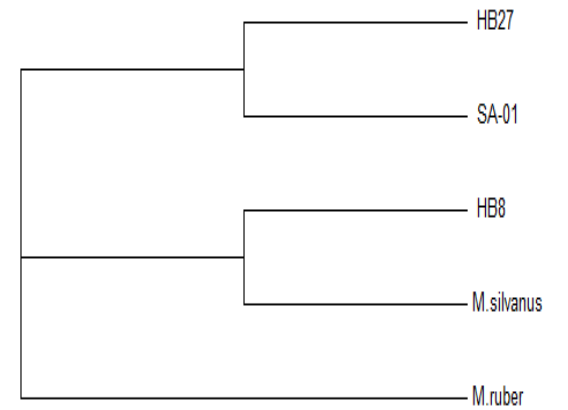


domain 1

Group 9

M.silvanus

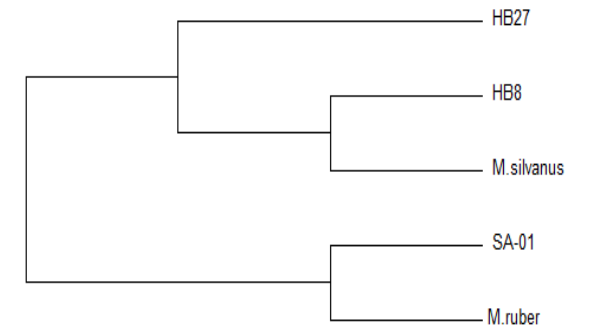
HB8 TSC_c03080 267629 268882 Dir cgtA Obg family GTPase



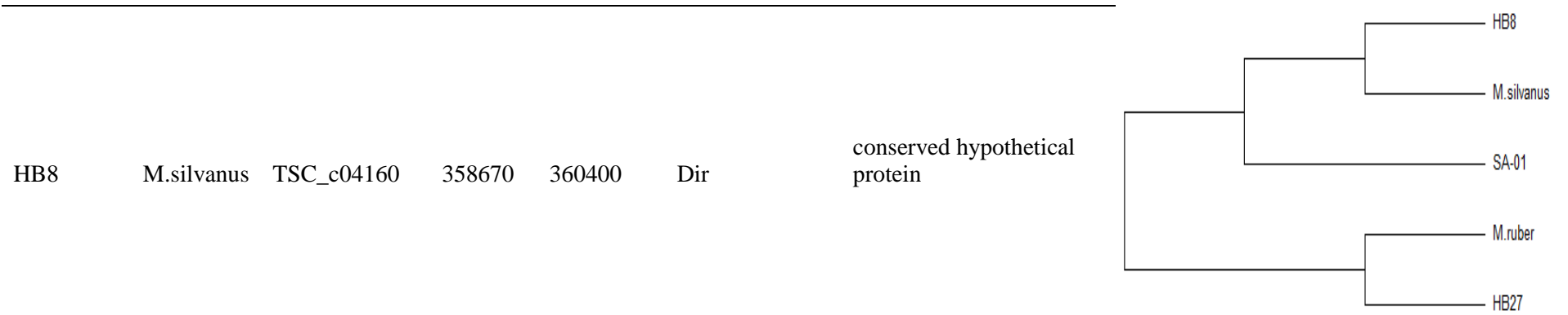
Group 10

M.silvanus

unknown TSC_c24070 2259618 2260424 Dir phnC phosphonate ABC transporter ATP-binding
148



protein



6.4 Appendix D

6.4.1 Thermostability analysis of orthologous sequences

Table 6.4.1: MFE values and resedu substitution statistics calculated for orthologous genes of *T. scotoductus* SA-01 and *T. thermophilus* HB27. Table entries are ordered by Δ MFE values (the difference between MFE values calculated for *T. thermophilus* HB27 and *T. scotoductus* SA-01 orthologous genes).

<i>T. scotoductus</i> SA-01			<i>T. thermophilus</i> HB27			Protein alignment		Δ MFE (kcal/Mol)	Total substitutions		Normalized substitutions [†]	
Gene ID	Length	MFE* (kcal/Mol)	Gene ID	Length	MFE* (kcal/Mol)	Length	Coverage		DNA	AMC	DNA	AMC
TSC_C18260	945	0.34	TTC0431	1017	0.53	341	95%	0.19	172	39	50.44	11.44
TSC_C18820	555	0.46	TTC0008	549	0.63	193	94%	0.17	112	33	58.03	17.10
TSC_C02960	1161	0.38	TTC0276	1095	0.55	441	82%	0.17	401	143	90.93	32.43
TSC_C22550	342	0.45	TTC1771	348	0.61	118	96%	0.16	75	25	63.56	21.19
TSC_C13260	1038	0.39	TTC0530	1050	0.54	359	96%	0.16	300	89	83.57	24.79
TSC_C11420	711	0.40	TTC0515	711	0.55	291	76%	0.15	242	94	83.16	32.30
TSC_C18350	537	0.32	TTC0423	537	0.47	216	78%	0.15	178	74	82.41	34.26
TSC_C08020	1044	0.43	TTC1829	1050	0.58	353	98%	0.14	167	42	47.31	11.90
TSC_C12740	402	0.45	TTC0626	417	0.59	147	91%	0.14	119	38	80.95	25.85
TSC_C08560	1140	0.49	TTC0368	1065	0.62	408	88%	0.14	295	94	72.30	23.04
TSC_C13580	960	0.43	TTC1833	960	0.56	319	100%	0.14	152	37	47.65	11.60
TSC_C01170	315	0.44	TTC1731	285	0.58	110	88%	0.14	79	27	71.82	24.55
TSC_C00780	306	0.39	TTC1805	303	0.52	111	89%	0.14	102	33	91.89	29.73
TSC_C21040	363	0.39	TTC1049	435	0.53	146	89%	0.14	67	16	45.89	10.96
TSC_C05390	387	0.37	TTC0125	393	0.50	135	95%	0.13	126	34	93.33	25.19
TSC_C00480	681	0.51	TTC1934	675	0.64	236	95%	0.13	178	61	75.42	25.85
TSC_C00380	1449	0.49	TTC1644	1440	0.62	530	90%	0.13	381	136	71.89	25.66
TSC_C07710	1665	0.32	TTC0857	1665	0.45	563	98%	0.13	447	123	79.40	21.85
TSC_C00510	819	0.45	TTC0630	924	0.58	336	84%	0.13	233	80	69.35	23.81

TSC_C13250	582	0.36	TTC0531	579	0.49	204	94%	0.13	153	41	75.00	20.10
TSC_C07880	615	0.49	TTC0706	618	0.61	214	95%	0.12	157	45	73.36	21.03
TSC_C14750	315	0.47	TTC0393	333	0.59	112	94%	0.12	61	14	54.46	12.50
TSC_C01450	807	0.50	TTC1633	807	0.62	278	96%	0.12	211	63	75.90	22.66
TSC_C20690	945	0.40	TTC1016	879	0.51	340	88%	0.12	245	79	72.06	23.24
TSC_C19830	1050	0.49	TTC1074	915	0.60	364	88%	0.12	246	93	67.58	25.55
TSC_C10340	444	0.46	TTC0587	435	0.57	152	95%	0.12	119	41	78.29	26.97
TSC_C15130	741	0.49	TTC0815	789	0.60	270	93%	0.11	158	47	58.52	17.41
TSC_C13360	1110	0.48	TTC0516	1110	0.59	373	99%	0.11	238	74	63.81	19.84
TSC_C21460	630	0.46	TTC1248	612	0.57	215	95%	0.11	145	43	67.44	20.00
TSC_C09160	279	0.43	TTC0348	276	0.54	92	98%	0.11	58	21	63.04	22.83
TSC_C23850	390	0.47	TTC1350	390	0.58	133	96%	0.11	80	29	60.15	21.80
TSC_C05570	897	0.51	TTC1063	888	0.63	317	93%	0.11	219	74	69.09	23.34
TSC_C24830	822	0.49	TTC1664	822	0.60	276	99%	0.11	166	42	60.14	15.22
TSC_C07380	1749	0.48	TTC0465	1758	0.60	615	94%	0.11	445	144	72.36	23.41
TSC_C20740	372	0.36	TTC0113	393	0.47	141	88%	0.11	123	44	87.23	31.21
TSC_C15120	762	0.51	TTC0814	753	0.63	271	92%	0.11	173	49	63.84	18.08
TSC_C20070	366	0.39	TTC1891	366	0.50	122	98%	0.11	59	22	48.36	18.03
TSC_C00730	462	0.46	TTC1590	468	0.57	158	97%	0.11	101	34	63.92	21.52
TSC_C00020	1116	0.46	TTC1609	1128	0.57	398	93%	0.11	257	85	64.57	21.36
TSC_C01040	762	0.48	TTC1720	786	0.59	269	95%	0.11	174	70	64.68	26.02
TSC_C09920	498	0.46	TTC0293	495	0.57	170	96%	0.10	123	41	72.35	24.12
TSC_C17230	924	0.49	TTC0233	1014	0.59	355	89%	0.10	199	71	56.06	20.00
TSC_C01750	165	0.26	TTC1735	165	0.36	54	98%	0.10	23	7	42.59	12.96
TSC_C14420	435	0.41	TTC0710	435	0.51	144	99%	0.10	60	13	41.67	9.03
TSC_C15540	360	0.46	TTC0353	372	0.56	125	96%	0.10	62	11	49.60	8.80
TSC_C08850	969	0.47	TTC0379	972	0.57	336	96%	0.10	246	84	73.21	25.00
TSC_C20670	624	0.43	TTC1014	624	0.53	210	98%	0.10	144	45	68.57	21.43
TSC_C24250	1236	0.51	TTC1554	1215	0.61	431	94%	0.10	261	72	60.56	16.71

TSC_C18460	336	0.47	TTC0941	333	0.57	120	91%	0.10	78	22	65.00	18.33
TSC_C21540	615	0.48	TTC1929	612	0.58	210	96%	0.10	147	42	70.00	20.00
TSC_C18010	552	0.47	TTC0737	546	0.57	189	96%	0.10	114	34	60.32	17.99
TSC_C08840	1191	0.51	TTC0378	1179	0.60	411	95%	0.10	272	95	66.18	23.11
TSC_C08690	372	0.45	TTC0916	354	0.55	135	87%	0.10	102	31	75.56	22.96
TSC_C14930	564	0.42	TTC0804	564	0.51	189	98%	0.10	112	35	59.26	18.52
TSC_C07920	1134	0.50	TTC0923	1119	0.60	385	97%	0.10	250	84	64.94	21.82
TSC_C24130	522	0.49	TTC1524	570	0.58	208	85%	0.10	158	63	75.96	30.29
TSC_C02930	909	0.37	TTC0528	873	0.47	350	81%	0.10	269	88	76.86	25.14
TSC_C00450	273	0.44	TTC1937	285	0.54	98	92%	0.10	78	31	79.59	31.63
TSC_C00680	489	0.50	TTC1587	492	0.59	167	97%	0.10	107	35	64.07	20.96
TSC_C13680	3033	0.51	TTC0543	3027	0.61	1083	93%	0.10	795	249	73.41	22.99
TSC_C01630	270	0.41	TTC1651	267	0.51	95	92%	0.10	45	17	47.37	17.89
TSC_C00760	561	0.47	TTC1803	564	0.57	191	97%	0.10	108	29	56.54	15.18
TSC_C23200	765	0.48	TTC1267	762	0.58	258	98%	0.10	146	43	56.59	16.67
TSC_C20680	624	0.40	TTC1015	582	0.49	211	94%	0.10	145	43	68.72	20.38
TSC_C00770	462	0.38	TTC1804	447	0.48	170	86%	0.10	140	47	82.35	27.65
TSC_C07690	702	0.50	TTC0863	657	0.60	241	93%	0.10	176	51	73.03	21.16
TSC_C01050	1023	0.51	TTC1721	1032	0.60	350	97%	0.10	230	69	65.71	19.71
TSC_C01580	726	0.50	TTC1624	729	0.60	245	98%	0.09	140	33	57.14	13.47
TSC_C21960	1098	0.52	TTC1200	1050	0.62	386	92%	0.09	254	66	65.80	17.10
TSC_C19730	1032	0.45	TTC0097	1032	0.55	345	99%	0.09	193	39	55.94	11.30
TSC_C04250	579	0.48	TTC1459	603	0.57	202	96%	0.09	121	29	59.90	14.36
TSC_C07760	360	0.34	TTC1284	336	0.43	137	80%	0.09	98	42	71.53	30.66
TSC_C02900	1629	0.50	TTC0478	1620	0.59	579	93%	0.09	427	158	73.75	27.29
TSC_C23810	1437	0.51	TTC1346	1425	0.61	505	94%	0.09	333	102	65.94	20.20
TSC_C06860	945	0.53	TTC0145	756	0.63	326	84%	0.09	193	55	59.20	16.87
TSC_C07100	453	0.47	TTC0894	402	0.56	165	83%	0.09	94	33	56.97	20.00
TSC_C22790	1011	0.47	TTC0030	1011	0.56	342	98%	0.09	217	73	63.45	21.35

TSC_C08890	351	0.47	TTC0383	375	0.56	128	93%	0.09	63	14	49.22	10.94
TSC_C23710	279	0.37	TTC1336	273	0.46	92	98%	0.09	35	9	38.04	9.78
TSC_C04890	231	0.40	TTC0993	231	0.49	81	92%	0.09	55	16	67.90	19.75
TSC_C21150	612	0.49	TTC1060	612	0.58	206	98%	0.09	98	25	47.57	12.14
TSC_C14180	591	0.46	TTC0214	591	0.55	199	98%	0.09	106	30	53.27	15.08
TSC_C07930	2931	0.52	TTC0922	2901	0.61	1084	88%	0.09	730	243	67.34	22.42
TSC_C14320	831	0.45	TTC0208	990	0.54	337	88%	0.09	137	40	40.65	11.87
TSC_C01650	618	0.49	TTC1653	597	0.58	215	93%	0.09	136	45	63.26	20.93
TSC_C23940	1746	0.53	TTC1686	1704	0.62	672	83%	0.09	472	166	70.24	24.70
TSC_C12950	1044	0.48	TTC0226	1044	0.58	353	98%	0.09	218	56	61.76	15.86
TSC_C05220	423	0.49	TTC0954	399	0.58	149	90%	0.09	107	39	71.81	26.17
TSC_C22940	369	0.44	TTC0005	339	0.53	124	93%	0.09	78	23	62.90	18.55
TSC_C19340	1512	0.47	TTC1187	1512	0.56	508	99%	0.09	231	48	45.47	9.45
TSC_C02880	414	0.51	TTC0741	402	0.60	146	91%	0.09	111	36	76.03	24.66
TSC_C24880	2127	0.51	TTC1942	2112	0.60	725	97%	0.09	500	132	68.97	18.21
TSC_C21360	348	0.50	TTC1239	339	0.59	127	87%	0.09	98	37	77.17	29.13
TSC_C06060	819	0.44	TTC1921	816	0.53	278	97%	0.09	158	49	56.83	17.63
TSC_C24480	696	0.52	TTC1965	696	0.61	235	98%	0.09	152	51	64.68	21.70
TSC_C07110	2238	0.48	TTC0893	2187	0.57	755	97%	0.09	431	135	57.09	17.88
TSC_C07810	453	0.43	TTC0196	453	0.51	151	99%	0.09	98	39	64.90	25.83
TSC_C06310	435	0.39	TTC1131	411	0.48	146	95%	0.09	68	20	46.58	13.70
TSC_C21100	327	0.38	TTC1055	327	0.47	108	99%	0.09	51	15	47.22	13.89
TSC_C06320	933	0.48	TTC1130	906	0.57	313	97%	0.09	217	70	69.33	22.36
TSC_C13840	621	0.46	TTC0668	579	0.55	211	94%	0.09	135	45	63.98	21.33
TSC_C18610	366	0.44	TTC1178	396	0.53	139	89%	0.09	105	28	75.54	20.14
TSC_C18340	1044	0.48	TTC0424	1041	0.57	353	98%	0.09	203	51	57.51	14.45
TSC_C15340	522	0.40	TTC0837	516	0.49	179	95%	0.09	156	45	87.15	25.14
TSC_C03890	303	0.42	TTC1490	303	0.51	100	99%	0.09	40	9	40.00	9.00
TSC_C00870	774	0.50	TTC1712	783	0.59	264	97%	0.09	166	50	62.88	18.94

TSC_C14480	858	0.47	TTC0716	720	0.56	290	89%	0.09	132	39	45.52	13.45
TSC_C23320	774	0.49	TTC1280	774	0.58	260	98%	0.09	146	38	56.15	14.62
TSC_C19330	468	0.47	TTC1186	468	0.56	156	99%	0.09	80	20	51.28	12.82
TSC_C14330	1110	0.51	TTC0207	1053	0.60	377	95%	0.09	221	62	58.62	16.45
TSC_C00880	2298	0.50	TTC1711	2238	0.59	775	97%	0.09	405	119	52.26	15.35
TSC_C11910	576	0.42	TTC0239	576	0.50	197	96%	0.09	122	40	61.93	20.30
TSC_C05380	1221	0.47	TTC0124	1221	0.56	414	98%	0.09	254	78	61.35	18.84
TSC_C07840	8085	0.56	TTC0200	8019	0.65	2864	93%	0.09	2154	751	75.21	26.22
TSC_C21400	600	0.48	TTC1243	597	0.57	209	94%	0.09	125	34	59.81	16.27
TSC_C04260	225	0.46	TTC1458	240	0.54	83	90%	0.08	40	14	48.19	16.87
TSC_C04300	798	0.48	TTC1454	804	0.57	276	96%	0.08	203	63	73.55	22.83
TSC_C21230	483	0.56	TTC1226	426	0.65	180	80%	0.08	84	29	46.67	16.11
TSC_C20980	567	0.48	TTC1043	570	0.56	195	96%	0.08	111	32	56.92	16.41
TSC_C20660	1134	0.45	TTC1013	1134	0.54	380	99%	0.08	253	83	66.58	21.84
TSC_C24960	498	0.48	TTC1599	495	0.56	169	97%	0.08	104	37	61.54	21.89
TSC_C11740	753	0.48	TTC0308	723	0.56	255	96%	0.08	170	53	66.67	20.78
TSC_C12700	1992	0.50	TTC0622	1929	0.58	685	95%	0.08	435	136	63.50	19.85
TSC_C01900	318	0.40	TTC1545	315	0.48	105	99%	0.08	39	9	37.14	8.57
TSC_C05690	258	0.37	TTC0076	249	0.45	87	95%	0.08	58	15	66.67	17.24
TSC_C14800	1194	0.53	TTC0329	1197	0.61	412	96%	0.08	289	93	70.15	22.57
TSC_C19560	999	0.47	TTC0115	999	0.55	335	99%	0.08	194	47	57.91	14.03
TSC_C23630	777	0.51	TTC1571	777	0.59	263	98%	0.08	169	52	64.26	19.77
TSC_C23920	237	0.45	TTC1689	237	0.54	78	99%	0.08	53	15	67.95	19.23
TSC_C19770	534	0.47	TTC0094	531	0.55	177	99%	0.08	102	31	57.63	17.51
TSC_C09700	588	0.38	TTC0767	591	0.47	196	99%	0.08	86	15	43.88	7.65
TSC_C05540	438	0.43	TTC1067	447	0.51	148	98%	0.08	81	24	54.73	16.22
TSC_C13240	846	0.44	TTC0532	933	0.52	311	94%	0.08	184	49	59.16	15.76
TSC_C18450	912	0.52	TTC0940	870	0.60	313	94%	0.08	190	51	60.70	16.29
TSC_C23760	927	0.51	TTC1340	783	0.60	316	88%	0.08	189	63	59.81	19.94

TSC_C20220	873	0.50	TTC0981	873	0.58	295	98%	0.08	172	54	58.31	18.31
TSC_C24550	648	0.48	TTC1956	684	0.56	235	93%	0.08	136	36	57.87	15.32
TSC_C06330	1146	0.48	TTC1129	1158	0.56	386	99%	0.08	247	66	63.99	17.10
TSC_C24870	1101	0.54	TTC1944	1086	0.62	384	94%	0.08	210	60	54.69	15.63
TSC_C08710	1146	0.46	TTC0914	1149	0.54	389	98%	0.08	240	69	61.70	17.74
TSC_C23280	570	0.46	TTC1276	585	0.54	209	90%	0.08	156	49	74.64	23.44
TSC_C09120	2583	0.54	TTC0344	2589	0.62	896	96%	0.08	571	141	63.73	15.74
TSC_C07870	993	0.47	TTC0707	993	0.55	333	99%	0.08	188	55	56.46	16.52
TSC_C02410	414	0.48	TTC1758	414	0.56	142	96%	0.08	77	19	54.23	13.38
TSC_C07860	1062	0.50	TTC0202	1059	0.59	357	98%	0.08	196	50	54.90	14.01
TSC_C04840	1251	0.50	TTC0998	1254	0.58	429	97%	0.08	262	85	61.07	19.81
TSC_C07490	309	0.41	TTC0880	306	0.49	103	98%	0.08	78	27	75.73	26.21
TSC_C21030	852	0.47	TTC1048	855	0.55	287	98%	0.08	187	61	65.16	21.25
TSC_C04520	687	0.49	TTC1977	687	0.57	228	100%	0.08	134	35	58.77	15.35
TSC_C00400	597	0.52	TTC1437	600	0.60	207	95%	0.08	91	16	43.96	7.73
TSC_C19580	315	0.43	TTC0118	309	0.51	107	95%	0.08	54	18	50.47	16.82
TSC_C21350	330	0.51	TTC1238	342	0.59	116	95%	0.08	68	32	58.62	27.59
TSC_C06880	1050	0.48	TTC0148	1050	0.56	354	98%	0.08	195	50	55.08	14.12
TSC_C17390	954	0.48	TTC0437	1083	0.56	396	83%	0.08	285	109	71.97	27.53
TSC_C14500	1059	0.48	TTC0718	1062	0.55	375	93%	0.08	247	78	65.87	20.80
TSC_C24850	705	0.51	TTC1666	705	0.59	249	93%	0.08	166	49	66.67	19.68
TSC_C19390	486	0.42	TTC1191	501	0.50	178	91%	0.08	101	35	56.74	19.66
TSC_C24180	1089	0.51	TTC1529	1065	0.59	365	98%	0.08	194	65	53.15	17.81
TSC_C21450	543	0.47	TTC1247	591	0.54	206	90%	0.08	118	40	57.28	19.42
TSC_C14870	687	0.51	TTC0298	687	0.59	231	98%	0.08	128	29	55.41	12.55
TSC_C18180	1332	0.50	TTC0889	1317	0.58	461	95%	0.08	259	69	56.18	14.97
TSC_C18240	1311	0.47	TTC0432	1311	0.55	443	98%	0.08	312	89	70.43	20.09
TSC_C22060	1260	0.51	TTC1212	1260	0.59	428	98%	0.08	255	79	59.58	18.46
TSC_C09800	330	0.47	TTC0757	339	0.55	117	93%	0.08	85	39	72.65	33.33

TSC_C20190	1326	0.53	TTC0088	1284	0.60	448	97%	0.08	246	52	54.91	11.61
TSC_C14000	819	0.50	TTC0256	741	0.57	281	91%	0.08	161	47	57.30	16.73
TSC_C17020	630	0.47	TTC0502	630	0.55	214	97%	0.08	141	44	65.89	20.56
TSC_C02190	1353	0.42	TTC1778	1353	0.50	451	100%	0.08	131	26	29.05	5.76
TSC_C21570	387	0.47	TTC1133	447	0.55	149	91%	0.08	89	26	59.73	17.45
TSC_C24310	1722	0.42	TTC1557	1671	0.50	599	94%	0.08	451	152	75.29	25.38
TSC_C04210	639	0.48	TTC1462	642	0.56	219	96%	0.08	144	44	65.75	20.09
TSC_C05660	1386	0.56	TTC0072	1344	0.63	492	91%	0.08	319	118	64.84	23.98
TSC_C08860	1269	0.44	TTC0380	1266	0.52	444	94%	0.08	319	102	71.85	22.97
TSC_C10580	1140	0.51	TTC0302	1200	0.59	403	96%	0.08	251	69	62.28	17.12
TSC_C21970	1455	0.51	TTC1201	1455	0.58	504	96%	0.08	348	100	69.05	19.84
TSC_C18620	633	0.53	TTC1179	540	0.61	217	88%	0.08	106	29	48.85	13.36
TSC_C24540	921	0.52	TTC1957	918	0.60	313	97%	0.08	165	41	52.72	13.10
TSC_C04530	675	0.47	TTC1978	672	0.55	226	98%	0.08	155	40	68.58	17.70
TSC_C19420	396	0.42	TTC1164	396	0.50	132	98%	0.08	84	27	63.64	20.45
TSC_C08670	828	0.48	TTC0918	783	0.55	283	94%	0.08	197	55	69.61	19.43
TSC_C15420	894	0.50	TTC0358	894	0.58	301	98%	0.08	188	59	62.46	19.60
TSC_C19940	318	0.40	TTC1900	318	0.47	107	97%	0.08	53	19	49.53	17.76
TSC_C00440	981	0.52	TTC1938	981	0.60	339	96%	0.08	189	58	55.75	17.11
TSC_C14840	909	0.53	TTC0305	909	0.60	308	98%	0.08	208	62	67.53	20.13
TSC_C01340	1560	0.48	TTC1787	1548	0.55	550	93%	0.08	373	117	67.82	21.27
TSC_C06120	1896	0.51	TTC0053	1887	0.59	649	97%	0.08	403	133	62.10	20.49
TSC_C09240	489	0.45	TTC0675	474	0.53	165	96%	0.08	67	18	40.61	10.91
TSC_C23860	750	0.54	TTC1695	750	0.61	267	92%	0.08	153	44	57.30	16.48
TSC_C13350	783	0.45	TTC0518	783	0.53	262	99%	0.08	134	33	51.15	12.60
TSC_C05420	1047	0.51	TTC0128	1044	0.58	359	96%	0.08	235	74	65.46	20.61
TSC_C07090	2652	0.53	TTC0895	2610	0.61	953	91%	0.08	618	228	64.85	23.92
TSC_C01270	1062	0.43	TTC1449	1080	0.51	372	95%	0.08	226	67	60.75	18.01
TSC_C14470	708	0.51	TTC0715	699	0.58	257	90%	0.08	186	68	72.37	26.46

TSC_C09480	693	0.49	TTC0410	705	0.56	243	95%	0.07	175	53	72.02	21.81
TSC_C06990	885	0.50	TTC0159	885	0.57	296	99%	0.07	133	34	44.93	11.49
TSC_C02870	1440	0.53	TTC0742	1434	0.61	513	92%	0.07	336	113	65.50	22.03
TSC_C15110	570	0.49	TTC0813	846	0.56	281	80%	0.07	106	25	37.72	8.90
TSC_C06380	1002	0.50	TTC1873	1005	0.58	341	97%	0.07	234	65	68.62	19.06
TSC_C04590	597	0.54	TTC1984	597	0.61	205	96%	0.07	109	32	53.17	15.61
TSC_C06350	810	0.46	TTC0979	816	0.54	273	99%	0.07	135	28	49.45	10.26
TSC_C23930	909	0.55	TTC1688	909	0.62	308	98%	0.07	159	48	51.62	15.58
TSC_C20480	558	0.47	TTC1105	612	0.54	206	93%	0.07	113	31	54.85	15.05
TSC_C06580	465	0.50	TTC1853	444	0.58	158	94%	0.07	87	33	55.06	20.89
TSC_C01870	222	0.33	TTC1548	225	0.41	75	97%	0.07	48	14	64.00	18.67
TSC_C18770	675	0.50	TTC0013	675	0.58	226	99%	0.07	133	39	58.85	17.26
TSC_C11890	444	0.37	TTC0237	456	0.45	151	98%	0.07	66	15	43.71	9.93
TSC_C00310	762	0.53	TTC0651	795	0.60	277	92%	0.07	187	66	67.51	23.83
TSC_C13690	327	0.46	TTC0542	327	0.54	108	99%	0.07	71	21	65.74	19.44
TSC_C15710	2487	0.53	TTC0479	2448	0.61	848	97%	0.07	559	172	65.92	20.28
TSC_C20830	759	0.48	TTC1026	765	0.56	262	96%	0.07	156	35	59.54	13.36
TSC_C08770	321	0.46	TTC0908	315	0.54	107	97%	0.07	59	19	55.14	17.76
TSC_C20770	378	0.42	TTC1022	378	0.49	127	98%	0.07	82	27	64.57	21.26
TSC_C11790	1731	0.45	TTC0313	1758	0.53	587	99%	0.07	292	58	49.74	9.88
TSC_C23570	1224	0.47	TTC1669	1221	0.54	438	92%	0.07	301	115	68.72	26.26
TSC_C22950	1353	0.55	TTC0004	1320	0.62	469	94%	0.07	330	107	70.36	22.81
TSC_C20570	1107	0.47	TTC1006	1083	0.55	374	97%	0.07	229	66	61.23	17.65
TSC_C17990	954	0.53	TTC0735	930	0.60	326	96%	0.07	191	58	58.59	17.79
TSC_C22930	1788	0.49	TTC0006	1776	0.56	606	98%	0.07	329	81	54.29	13.37
TSC_C05350	1122	0.48	TTC0123	1098	0.55	383	96%	0.07	244	62	63.71	16.19
TSC_C07900	369	0.48	TTC0925	369	0.55	125	97%	0.07	88	28	70.40	22.40
TSC_C05610	1842	0.35	TTC1430	1836	0.42	644	95%	0.07	535	142	83.07	22.05
TSC_C01180	258	0.44	TTC1732	258	0.51	88	95%	0.07	72	24	81.82	27.27

TSC_C17820	855	0.46	TTC0704	855	0.54	289	98%	0.07	176	48	60.90	16.61
TSC_C19470	1545	0.51	TTC1176	1524	0.59	530	96%	0.07	324	90	61.13	16.98
TSC_C00200	978	0.51	TTC1434	921	0.58	368	83%	0.07	232	72	63.04	19.57
TSC_C23870	384	0.49	TTC1694	369	0.57	127	97%	0.07	65	16	51.18	12.60
TSC_C24190	1116	0.51	TTC1530	1023	0.58	396	88%	0.07	208	62	52.53	15.66
TSC_C24950	819	0.49	TTC1598	795	0.56	278	96%	0.07	144	32	51.80	11.51
TSC_C10960	1170	0.46	TTC1108	1161	0.53	476	77%	0.07	347	142	72.90	29.83
TSC_C15300	573	0.49	TTC0833	579	0.56	204	93%	0.07	168	56	82.35	27.45
TSC_C08080	1278	0.53	TTC0173	1272	0.60	432	98%	0.07	221	58	51.16	13.43
TSC_C19150	1038	0.51	TTC0037	1032	0.58	351	98%	0.07	221	57	62.96	16.24
TSC_C05560	954	0.48	TTC1064	954	0.55	320	99%	0.07	194	56	60.63	17.50
TSC_C08820	753	0.53	TTC0376	747	0.60	257	96%	0.07	161	56	62.65	21.79
TSC_C05120	441	0.49	TTC0963	438	0.56	148	98%	0.07	86	21	58.11	14.19
TSC_C06200	927	0.52	TTC0063	921	0.59	309	99%	0.07	147	36	47.57	11.65
TSC_C18930	300	0.47	TTC1086	300	0.54	100	98%	0.07	55	19	55.00	19.00
TSC_C08440	534	0.41	TTC1413	561	0.48	204	87%	0.07	155	50	75.98	24.51
TSC_C22110	804	0.51	TTC1294	771	0.58	288	90%	0.07	176	52	61.11	18.06
TSC_C21990	825	0.47	TTC1203	819	0.54	279	97%	0.07	156	44	55.91	15.77
TSC_C21210	2619	0.54	TTC0638	2574	0.61	919	94%	0.07	561	185	61.04	20.13
TSC_C09010	750	0.44	TTC0333	756	0.51	253	98%	0.07	129	40	50.99	15.81
TSC_C15370	675	0.43	TTC0364	675	0.50	229	97%	0.07	140	37	61.14	16.16
TSC_C02100	291	0.46	TTC1442	327	0.53	110	91%	0.07	48	15	43.64	13.64
TSC_C23530	723	0.51	TTC1510	723	0.58	242	99%	0.07	132	43	54.55	17.77
TSC_C07990	2328	0.48	TTC0902	2313	0.55	781	99%	0.07	372	113	47.63	14.47
TSC_C08570	600	0.51	TTC0369	567	0.58	208	92%	0.07	126	42	60.58	20.19
TSC_C01190	858	0.50	TTC1800	858	0.57	291	98%	0.07	152	47	52.23	16.15
TSC_C11730	1050	0.45	TTC0307	1050	0.52	354	98%	0.07	210	55	59.32	15.54
TSC_C09490	411	0.46	TTC0409	411	0.53	138	98%	0.07	106	28	76.81	20.29
TSC_C20840	1458	0.53	TTC1027	1449	0.60	498	97%	0.07	288	67	57.83	13.45

TSC_C15200	1125	0.48	TTC0822	1107	0.55	377	98%	0.07	167	35	44.30	9.28
TSC_C14190	1059	0.49	TTC0213	1059	0.56	358	98%	0.07	204	49	56.98	13.69
TSC_C19050	381	0.43	TTC1097	381	0.50	127	98%	0.07	52	16	40.94	12.60
TSC_C22250	483	0.50	TTC1309	453	0.57	160	96%	0.07	58	12	36.25	7.50
TSC_C06440	1086	0.51	TTC1865	1086	0.58	369	98%	0.07	219	58	59.35	15.72
TSC_C23400	1713	0.46	TTC1283	1713	0.53	581	98%	0.07	353	104	60.76	17.90
TSC_C17250	1356	0.56	TTC0230	1374	0.62	467	97%	0.07	269	80	57.60	17.13
TSC_C04850	684	0.49	TTC0997	684	0.55	227	100%	0.07	114	26	50.22	11.45
TSC_C07270	444	0.38	TTC0454	441	0.45	153	95%	0.07	92	38	60.13	24.84
TSC_C03380	561	0.50	TTC1819	570	0.57	190	98%	0.07	92	26	48.42	13.68
TSC_C24320	915	0.55	TTC1663	918	0.62	315	96%	0.07	206	66	65.40	20.95
TSC_C01280	591	0.45	TTC1448	585	0.52	205	94%	0.07	118	41	57.56	20.00
TSC_C19970	669	0.44	TTC1897	663	0.51	228	96%	0.07	121	27	53.07	11.84
TSC_C02940	1113	0.43	TTC0527	1212	0.50	484	75%	0.07	368	140	76.03	28.93
TSC_C21470	528	0.45	TTC1249	528	0.52	178	98%	0.07	87	21	48.88	11.80
TSC_C12920	1671	0.52	TTC0228	1662	0.59	569	97%	0.07	334	91	58.70	15.99
TSC_C09720	1314	0.49	TTC0765	1326	0.56	444	99%	0.07	241	58	54.28	13.06
TSC_C14520	1341	0.51	TTC0720	1341	0.58	464	96%	0.07	322	87	69.40	18.75
TSC_C20040	1215	0.51	TTC1894	1212	0.58	408	99%	0.07	213	50	52.21	12.25
TSC_C19280	969	0.47	TTC1223	960	0.54	323	99%	0.07	207	51	64.09	15.79
TSC_C24840	1254	0.52	TTC1665	1227	0.58	434	95%	0.07	266	86	61.29	19.82
TSC_C04830	462	0.45	TTC0066	486	0.51	166	94%	0.07	111	38	66.87	22.89
TSC_C25040	753	0.50	TTC1605	750	0.57	250	99%	0.07	128	35	51.20	14.00
TSC_C02260	816	0.52	TTC1745	825	0.59	285	95%	0.07	164	39	57.54	13.68
TSC_C05670	591	0.48	TTC0074	570	0.55	200	96%	0.07	117	42	58.50	21.00
TSC_C21290	1752	0.48	TTC1232	1716	0.55	605	95%	0.07	361	115	59.67	19.01
TSC_C03590	987	0.53	TTC1384	981	0.59	333	98%	0.07	193	49	57.96	14.71
TSC_C09330	996	0.43	TTC0684	996	0.50	333	99%	0.07	177	36	53.15	10.81
TSC_C17070	438	0.45	TTC0497	444	0.52	147	99%	0.07	67	16	45.58	10.88

TSC_C05620	915	0.49	TTC0068	915	0.56	305	99%	0.07	129	29	42.30	9.51
TSC_C07910	1017	0.48	TTC0924	1011	0.55	357	94%	0.07	237	69	66.39	19.33
TSC_C23900	717	0.50	TTC1691	717	0.56	242	98%	0.07	125	33	51.65	13.64
TSC_C05900	984	0.50	TTC1132	903	0.57	333	93%	0.07	172	42	51.65	12.61
TSC_C09710	492	0.43	TTC0766	492	0.50	164	99%	0.07	88	25	53.66	15.24
TSC_C19860	495	0.51	TTC1902	504	0.58	177	92%	0.07	104	31	58.76	17.51
TSC_C17790	1113	0.49	TTC0411	1092	0.56	374	98%	0.07	201	59	53.74	15.78
TSC_C07440	2532	0.53	TTC0885	2538	0.59	872	97%	0.07	496	162	56.88	18.58
TSC_C19820	330	0.43	TTC0134	330	0.50	111	97%	0.07	67	18	60.36	16.22
TSC_C14490	1254	0.49	TTC0717	1254	0.56	425	98%	0.07	248	64	58.35	15.06
TSC_C20370	447	0.36	TTC1115	447	0.42	148	99%	0.07	71	20	47.97	13.51
TSC_C06980	951	0.54	TTC0158	966	0.61	324	98%	0.07	183	48	56.48	14.81
TSC_C20210	774	0.49	TTC0980	765	0.56	266	96%	0.07	197	67	74.06	25.19
TSC_C01640	801	0.47	TTC1652	804	0.53	274	97%	0.07	145	40	52.92	14.60
TSC_C18290	330	0.44	TTC0429	330	0.50	109	99%	0.07	79	27	72.48	24.77
TSC_C06090	345	0.42	TTC1924	339	0.49	120	93%	0.07	50	18	41.67	15.00
TSC_C20910	981	0.52	TTC1035	981	0.58	328	99%	0.07	201	56	61.28	17.07
TSC_C24860	321	0.41	TTC1945	321	0.48	108	97%	0.07	67	15	62.04	13.89
TSC_C19060	627	0.50	TTC1098	627	0.56	211	98%	0.07	115	35	54.50	16.59
TSC_C14890	597	0.50	TTC0296	603	0.57	202	98%	0.07	141	42	69.80	20.79
TSC_C21070	1662	0.47	TTC1052	1722	0.53	576	98%	0.07	287	83	49.83	14.41
TSC_C21190	432	0.39	TTC1101	492	0.45	164	92%	0.07	57	17	34.76	10.37
TSC_C09890	849	0.51	TTC0747	852	0.58	294	96%	0.07	198	66	67.35	22.45
TSC_C19670	627	0.50	TTC0103	627	0.56	209	99%	0.07	93	23	44.50	11.00
TSC_C09370	795	0.53	TTC0688	792	0.59	264	99%	0.07	163	50	61.74	18.94
TSC_C00100	414	0.43	TTC1798	414	0.49	137	99%	0.07	58	10	42.34	7.30
TSC_C15240	789	0.55	TTC0826	789	0.61	270	97%	0.07	173	49	64.07	18.15
TSC_C06610	663	0.42	TTC1851	663	0.48	220	100%	0.07	100	29	45.45	13.18
TSC_C11920	1092	0.44	TTC0616	1092	0.50	365	99%	0.07	176	31	48.22	8.49

TSC_C09040	906	0.49	TTC0336	900	0.55	303	99%	0.07	149	37	49.17	12.21
TSC_C12720	1539	0.54	TTC0624	1485	0.61	523	96%	0.07	365	84	69.79	16.06
TSC_C14710	909	0.46	TTC0406	912	0.52	307	98%	0.07	165	41	53.75	13.36
TSC_C24900	660	0.52	TTC1940	660	0.58	224	97%	0.06	160	44	71.43	19.64
TSC_C19010	435	0.41	TTC1093	492	0.48	164	93%	0.06	52	14	31.71	8.54
TSC_C02640	576	0.51	TTC1696	576	0.58	192	99%	0.06	105	38	54.69	19.79
TSC_C22770	1035	0.53	TTC0032	1035	0.59	344	100%	0.06	146	37	42.44	10.76
TSC_C13160	1023	0.49	TTC0538	1026	0.56	343	99%	0.06	177	35	51.60	10.20
TSC_C06420	576	0.42	TTC1867	585	0.48	194	99%	0.06	92	19	47.42	9.79
TSC_C19130	786	0.49	TTC0039	855	0.56	287	94%	0.06	115	23	40.07	8.01
TSC_C22710	1188	0.44	TTC1219	1188	0.50	397	99%	0.06	129	20	32.49	5.04
TSC_C04820	210	0.44	TTC0065	231	0.51	78	91%	0.06	36	11	46.15	14.10
TSC_C14880	765	0.51	TTC0297	765	0.57	262	96%	0.06	157	44	59.92	16.79
TSC_C18910	756	0.44	TTC1084	756	0.50	254	98%	0.06	161	46	63.39	18.11
TSC_C14860	894	0.43	TTC0300	903	0.50	303	98%	0.06	152	52	50.17	17.16
TSC_C09810	456	0.45	TTC0756	456	0.51	152	99%	0.06	90	28	59.21	18.42
TSC_C08630	1092	0.47	TTC0921	1029	0.54	367	96%	0.06	205	64	55.86	17.44
TSC_C07150	594	0.47	TTC0439	588	0.53	202	96%	0.06	111	26	54.95	12.87
TSC_C06520	483	0.47	TTC1859	465	0.53	167	93%	0.06	97	22	58.08	13.17
TSC_C07320	891	0.55	TTC0459	855	0.62	305	95%	0.06	183	50	60.00	16.39
TSC_C22000	2784	0.54	TTC1204	2751	0.60	939	98%	0.06	486	130	51.76	13.84
TSC_C20520	570	0.46	TTC1002	564	0.52	197	95%	0.06	135	44	68.53	22.34
TSC_C17100	279	0.48	TTC0493	279	0.54	93	98%	0.06	41	13	44.09	13.98
TSC_C10140	393	0.50	TTC0658	336	0.57	138	85%	0.06	87	36	63.04	26.09
TSC_C17940	801	0.41	TTC0731	813	0.48	277	96%	0.06	173	40	62.45	14.44
TSC_C22650	1389	0.48	TTC1493	1389	0.54	462	100%	0.06	194	35	41.99	7.58
TSC_C03360	600	0.45	TTC1817	567	0.52	207	93%	0.06	111	42	53.62	20.29
TSC_C13660	1146	0.54	TTC0545	1155	0.61	405	94%	0.06	303	105	74.81	25.93
TSC_C19450	609	0.48	TTC1166	630	0.54	221	92%	0.06	159	51	71.95	23.08

TSC_C18880	681	0.45	TTC0007	651	0.52	229	96%	0.06	124	35	54.15	15.28
TSC_C12080	1047	0.44	TTC0136	1056	0.51	370	94%	0.06	303	94	81.89	25.41
TSC_C20030	1566	0.52	TTC1895	1554	0.58	538	96%	0.06	301	89	55.95	16.54
TSC_C23460	453	0.45	TTC1502	453	0.51	151	99%	0.06	85	29	56.29	19.21
TSC_C11780	765	0.46	TTC0312	762	0.52	259	97%	0.06	171	56	66.02	21.62
TSC_C09600	2439	0.49	TTC0776	2430	0.55	831	97%	0.06	505	133	60.77	16.00
TSC_C19990	1956	0.47	TTC1896	1956	0.53	657	99%	0.06	324	77	49.32	11.72
TSC_C22570	939	0.49	TTC1769	939	0.56	313	99%	0.06	173	36	55.27	11.50
TSC_C10460	504	0.51	TTC0595	450	0.58	168	93%	0.06	94	25	55.95	14.88
TSC_C09180	426	0.48	TTC0350	474	0.54	158	93%	0.06	74	16	46.84	10.13
TSC_C03530	240	0.37	TTC1395	240	0.43	79	99%	0.06	50	14	63.29	17.72
TSC_C15580	672	0.45	TTC0701	672	0.51	227	98%	0.06	98	22	43.17	9.69
TSC_C04290	1503	0.49	TTC1455	1491	0.55	525	94%	0.06	284	77	54.10	14.67
TSC_C23070	261	0.44	TTC0015	246	0.50	89	92%	0.06	60	20	67.42	22.47
TSC_C09970	1395	0.47	TTC0268	1509	0.53	503	96%	0.06	238	49	47.32	9.74
TSC_C23450	1689	0.48	TTC1501	1692	0.54	567	99%	0.06	237	73	41.80	12.87
TSC_C04610	993	0.50	TTC1986	993	0.56	334	98%	0.06	239	78	71.56	23.35
TSC_C11820	786	0.49	TTC0317	807	0.56	271	97%	0.06	146	32	53.87	11.81
TSC_C15210	684	0.50	TTC0823	663	0.56	229	97%	0.06	136	37	59.39	16.16
TSC_C14270	849	0.44	TTC0212	843	0.50	298	94%	0.06	196	61	65.77	20.47
TSC_C04960	984	0.58	TTC0986	999	0.64	342	96%	0.06	195	57	57.02	16.67
TSC_C10320	843	0.45	TTC0585	843	0.52	284	98%	0.06	167	50	58.80	17.61
TSC_C23640	726	0.45	TTC1570	732	0.51	247	98%	0.06	130	42	52.63	17.00
TSC_C23700	477	0.48	TTC1335	477	0.54	160	98%	0.06	112	33	70.00	20.63
TSC_C05450	705	0.51	TTC0131	825	0.58	287	87%	0.06	160	48	55.75	16.72
TSC_C24100	684	0.54	TTC1522	678	0.60	234	96%	0.06	157	40	67.09	17.09
TSC_C03370	495	0.47	TTC1818	495	0.53	164	99%	0.06	56	8	34.15	4.88
TSC_C24590	1560	0.50	TTC1951	1536	0.56	526	98%	0.06	257	60	48.86	11.41
TSC_C09320	417	0.48	TTC0683	420	0.54	139	99%	0.06	64	12	46.04	8.63

TSC_C18650	1779	0.49	TTC1182	1893	0.55	640	95%	0.06	307	82	47.97	12.81
TSC_C01130	294	0.43	TTC1148	273	0.49	103	89%	0.06	81	27	78.64	26.21
TSC_C22800	1044	0.49	TTC0029	1044	0.55	353	98%	0.06	186	57	52.69	16.15
TSC_C24300	636	0.46	TTC1556	597	0.52	229	88%	0.06	181	63	79.04	27.51
TSC_C20390	474	0.48	TTC1112	474	0.54	157	99%	0.06	74	20	47.13	12.74
TSC_C07600	534	0.45	TTC0869	537	0.51	184	96%	0.06	123	42	66.85	22.83
TSC_C22960	543	0.48	TTC0003	543	0.54	184	97%	0.06	91	18	49.46	9.78
TSC_C08910	855	0.48	TTC0385	852	0.54	285	99%	0.06	161	43	56.49	15.09
TSC_C08250	429	0.51	TTC0180	429	0.57	146	97%	0.06	93	33	63.70	22.60
TSC_C17050	1200	0.51	TTC0499	1197	0.56	407	98%	0.06	254	63	62.41	15.48
TSC_C01560	483	0.45	TTC1623	486	0.51	165	97%	0.06	74	23	44.85	13.94
TSC_C20240	636	0.45	TTC0983	642	0.51	215	98%	0.06	112	30	52.09	13.95
TSC_C13640	711	0.53	TTC0547	708	0.59	251	93%	0.06	164	52	65.34	20.72
TSC_C14360	909	0.52	TTC0204	909	0.58	304	99%	0.06	169	47	55.59	15.46
TSC_C21420	450	0.50	TTC1245	537	0.56	183	88%	0.06	101	30	55.19	16.39
TSC_C09380	579	0.47	TTC0689	576	0.53	208	91%	0.06	103	29	49.52	13.94
TSC_C12730	2034	0.48	TTC0625	1869	0.54	701	92%	0.06	471	155	67.19	22.11
TSC_C20530	807	0.47	TTC1003	804	0.52	272	98%	0.06	143	37	52.57	13.60
TSC_C01660	552	0.47	TTC1654	558	0.53	195	94%	0.06	134	40	68.72	20.51
TSC_C03240	1617	0.52	TTC1407	1410	0.58	544	92%	0.06	306	98	56.25	18.01
TSC_C13670	486	0.42	TTC0544	498	0.48	181	88%	0.06	124	50	68.51	27.62
TSC_C24110	1248	0.50	TTC1523	1266	0.56	424	98%	0.06	179	36	42.22	8.49
TSC_C24780	1179	0.49	TTC1558	1179	0.55	405	96%	0.06	283	102	69.88	25.19
TSC_C10970	420	0.48	TTC0539	414	0.53	162	82%	0.06	132	50	81.48	30.86
TSC_C00140	834	0.52	TTC1794	840	0.58	293	94%	0.06	248	82	84.64	27.99
TSC_C23020	1140	0.47	TTC1933	1119	0.53	389	96%	0.06	251	77	64.52	19.79
TSC_C02130	402	0.43	TTC1443	387	0.49	136	95%	0.06	64	22	47.06	16.18
TSC_C07040	1182	0.50	TTC0165	1128	0.56	403	95%	0.06	231	66	57.32	16.38
TSC_C11760	669	0.39	TTC0310	669	0.45	224	99%	0.06	132	38	58.93	16.96

TSC_C01300	1296	0.50	TTC1446	1305	0.56	435	99%	0.06	216	50	49.66	11.49
TSC_C02140	1485	0.53	TTC1785	1485	0.59	502	98%	0.06	212	55	42.23	10.96
TSC_C00060	1470	0.51	TTC1427	1464	0.56	504	97%	0.06	349	111	69.25	22.02
TSC_C01150	1398	0.51	TTC1729	1578	0.57	537	91%	0.06	266	79	49.53	14.71
TSC_C23830	813	0.53	TTC1348	843	0.59	295	92%	0.06	170	56	57.63	18.98
TSC_C20750	1368	0.52	TTC1021	1344	0.58	461	98%	0.06	266	74	57.70	16.05
TSC_C13460	1299	0.50	TTC0562	1302	0.56	439	98%	0.06	271	62	61.73	14.12
TSC_C01160	444	0.37	TTC1730	444	0.43	148	99%	0.06	77	17	52.03	11.49
TSC_C22840	402	0.51	TTC0024	393	0.56	136	96%	0.06	64	13	47.06	9.56
TSC_C23290	1893	0.46	TTC1277	1893	0.52	637	99%	0.06	263	67	41.29	10.52
TSC_C24340	819	0.56	TTC1661	810	0.61	295	91%	0.06	188	60	63.73	20.34
TSC_C23500	276	0.50	TTC1506	258	0.56	96	90%	0.06	58	25	60.42	26.04
TSC_C02070	1077	0.53	TTC1441	1068	0.58	377	94%	0.06	226	69	59.95	18.30
TSC_C13880	219	0.41	TTC0672	219	0.47	72	99%	0.06	51	16	70.83	22.22
TSC_C07120	624	0.44	TTC0436	618	0.50	219	93%	0.06	136	50	62.10	22.83
TSC_C04080	726	0.50	TTC1474	714	0.56	244	98%	0.06	154	43	63.11	17.62
TSC_C18640	963	0.51	TTC1181	843	0.57	328	90%	0.06	188	59	57.32	17.99
TSC_C08760	972	0.53	TTC0909	972	0.58	325	99%	0.06	192	54	59.08	16.62
TSC_C19850	639	0.44	TTC1903	618	0.50	218	95%	0.06	131	41	60.09	18.81
TSC_C15690	660	0.40	TTC0481	660	0.46	222	98%	0.06	159	38	71.62	17.12
TSC_C20320	390	0.50	TTC1122	390	0.55	132	97%	0.06	70	18	53.03	13.64
TSC_C06140	717	0.53	TTC0055	717	0.59	243	97%	0.06	160	48	65.84	19.75
TSC_C15610	1035	0.49	TTC0703	1035	0.55	349	98%	0.06	202	62	57.88	17.77
TSC_C14130	1551	0.47	TTC0243	1551	0.53	522	99%	0.06	326	98	62.45	18.77
TSC_C13650	435	0.38	TTC0546	435	0.44	146	98%	0.06	61	11	41.78	7.53
TSC_C15280	438	0.43	TTC0831	438	0.48	146	99%	0.06	102	27	69.86	18.49
TSC_C01060	1173	0.50	TTC1722	1173	0.56	401	97%	0.06	208	51	51.87	12.72
TSC_C15630	816	0.51	TTC0482	813	0.56	276	98%	0.06	142	39	51.45	14.13
TSC_C15520	567	0.48	TTC0355	576	0.54	194	97%	0.06	85	16	43.81	8.25

TSC_C17880	684	0.49	TTC0932	684	0.55	231	98%	0.06	119	25	51.52	10.82
TSC_C21900	477	0.44	TTC1194	495	0.50	165	97%	0.06	77	25	46.67	15.15
TSC_C05460	699	0.45	TTC0132	696	0.51	237	97%	0.06	58	13	24.47	5.49
TSC_C14150	849	0.43	TTC0240	858	0.49	288	98%	0.06	147	32	51.04	11.11
TSC_C13040	942	0.49	TTC0781	942	0.54	317	98%	0.06	163	37	51.42	11.67
TSC_C04900	423	0.42	TTC0992	426	0.48	141	99%	0.06	96	22	68.09	15.60
TSC_C23620	1218	0.49	TTC1572	1197	0.55	409	98%	0.06	212	75	51.83	18.34
TSC_C22120	1383	0.55	TTC1295	1362	0.61	482	94%	0.06	352	110	73.03	22.82
TSC_C04500	2088	0.53	TTC1975	2097	0.59	710	98%	0.06	374	112	52.68	15.77
TSC_C21120	342	0.46	TTC1057	462	0.52	155	83%	0.06	62	20	40.00	12.90
TSC_C02860	636	0.46	TTC0743	630	0.52	219	95%	0.06	127	44	57.99	20.09
TSC_C22780	411	0.49	TTC0031	417	0.54	140	97%	0.06	77	22	55.00	15.71
TSC_C18000	840	0.47	TTC0736	624	0.52	280	84%	0.06	101	28	36.07	10.00
TSC_C02580	1263	0.49	TTC0645	1257	0.54	453	92%	0.06	387	136	85.43	30.02
TSC_C21320	318	0.42	TTC1235	318	0.47	105	99%	0.06	48	13	45.71	12.38
TSC_C23270	717	0.51	TTC1275	717	0.57	242	98%	0.06	148	44	61.16	18.18
TSC_C18080	2604	0.49	TTC0474	2604	0.54	907	95%	0.06	480	144	52.92	15.88
TSC_C20720	555	0.45	TTC1019	555	0.51	185	99%	0.06	98	30	52.97	16.22
TSC_C20730	1035	0.50	TTC1020	1047	0.56	355	97%	0.06	211	62	59.44	17.46
TSC_C06540	375	0.46	TTC1857	375	0.51	124	99%	0.05	67	18	54.03	14.52
TSC_C23690	1116	0.51	TTC1334	1125	0.57	378	98%	0.05	192	45	50.79	11.90
TSC_C22580	786	0.49	TTC1768	795	0.55	267	98%	0.05	154	41	57.68	15.36
TSC_C21710	639	0.54	TTC1143	678	0.59	233	93%	0.05	134	36	57.51	15.45
TSC_C20250	294	0.37	TTC0984	294	0.42	97	99%	0.05	34	7	35.05	7.22
TSC_C07330	1248	0.53	TTC0460	1386	0.58	464	94%	0.05	202	47	43.53	10.13
TSC_C15090	876	0.50	TTC0812	858	0.56	306	93%	0.05	212	55	69.28	17.97
TSC_C19880	855	0.52	TTC1901	855	0.57	285	99%	0.05	113	23	39.65	8.07
TSC_C07610	540	0.44	TTC0868	534	0.49	185	96%	0.05	132	45	71.35	24.32
TSC_C18050	633	0.46	TTC0477	633	0.52	210	100%	0.05	98	23	46.67	10.95

TSC_C08940	984	0.51	TTC0388	975	0.57	339	96%	0.05	190	50	56.05	14.75
TSC_C20450	1557	0.54	TTC1107	1608	0.60	555	94%	0.05	299	84	53.87	15.14
TSC_C19490	210	0.44	TTC0943	207	0.49	69	98%	0.05	41	14	59.42	20.29
TSC_C22860	942	0.52	TTC0021	765	0.57	315	89%	0.05	131	26	41.59	8.25
TSC_C12220	1533	0.55	TTC0619	1482	0.60	535	93%	0.05	331	89	61.87	16.64
TSC_C20640	519	0.45	TTC1011	492	0.51	180	92%	0.05	121	25	67.22	13.89
TSC_C19320	666	0.56	TTC1185	663	0.61	225	98%	0.05	125	31	55.56	13.78
TSC_C01610	333	0.45	TTC1649	333	0.51	112	97%	0.05	51	17	45.54	15.18
TSC_C03790	621	0.49	TTC1366	627	0.54	223	92%	0.05	179	65	80.27	29.15
TSC_C08920	1932	0.48	TTC0386	1926	0.54	678	94%	0.05	514	176	75.81	25.96
TSC_C22820	1002	0.47	TTC0027	990	0.52	356	92%	0.05	259	92	72.75	25.84
TSC_C19160	1353	0.54	TTC0036	1374	0.59	465	97%	0.05	267	82	57.42	17.63
TSC_C00650	471	0.49	TTC1584	471	0.55	159	97%	0.05	71	23	44.65	14.47
TSC_C19080	1164	0.52	TTC0087	1125	0.58	393	96%	0.05	200	48	50.89	12.21
TSC_C23540	624	0.49	TTC1511	630	0.54	213	97%	0.05	163	46	76.53	21.60
TSC_C19170	1854	0.47	TTC0035	1857	0.52	621	99%	0.05	263	58	42.35	9.34
TSC_C06740	279	0.48	TTC0654	285	0.54	103	88%	0.05	58	23	56.31	22.33
TSC_C23190	1053	0.54	TTC1266	1053	0.59	356	98%	0.05	174	52	48.88	14.61
TSC_C09460	606	0.46	TTC0557	612	0.51	208	97%	0.05	108	35	51.92	16.83
TSC_C18320	1281	0.48	TTC0426	1281	0.53	429	99%	0.05	263	61	61.31	14.22
TSC_C18300	546	0.44	TTC0428	546	0.50	181	99%	0.05	77	18	42.54	9.94
TSC_C05510	708	0.52	TTC1070	708	0.57	238	98%	0.05	148	44	62.18	18.49
TSC_C22560	1110	0.49	TTC1770	1110	0.55	371	99%	0.05	146	22	39.35	5.93
TSC_C02280	675	0.48	TTC1746	654	0.54	226	97%	0.05	99	24	43.81	10.62
TSC_C07560	1059	0.46	TTC0873	1056	0.51	356	98%	0.05	176	48	49.44	13.48
TSC_C00230	1011	0.49	TTC1612	987	0.54	354	93%	0.05	212	66	59.89	18.64
TSC_C10590	2589	0.53	TTC0304	2598	0.58	875	99%	0.05	444	115	50.74	13.14
TSC_C14140	372	0.42	TTC0241	351	0.47	123	96%	0.05	49	12	39.84	9.76
TSC_C23030	2550	0.50	TTC1931	2475	0.55	879	95%	0.05	441	116	50.17	13.20

TSC_C09360	1269	0.50	TTC0687	1302	0.55	436	98%	0.05	235	53	53.90	12.16
TSC_C08000	1215	0.53	TTC0901	1212	0.58	409	98%	0.05	211	53	51.59	12.96
TSC_C18800	1194	0.48	TTC0010	1194	0.53	405	98%	0.05	177	38	43.70	9.38
TSC_C24910	1062	0.48	TTC1594	1053	0.53	356	98%	0.05	163	44	45.79	12.36
TSC_C06560	213	0.38	TTC1855	234	0.44	78	93%	0.05	30	7	38.46	8.97
TSC_C13410	513	0.52	TTC0559	558	0.57	187	94%	0.05	84	24	44.92	12.83
TSC_C24610	351	0.44	TTC1949	351	0.49	116	99%	0.05	45	9	38.79	7.76
TSC_C07750	309	0.48	TTC0192	309	0.53	103	98%	0.05	57	16	55.34	15.53
TSC_C03930	2763	0.52	TTC1486	2757	0.58	940	98%	0.05	539	157	57.34	16.70
TSC_C14430	861	0.48	TTC0711	858	0.54	291	98%	0.05	190	60	65.29	20.62
TSC_C04550	1320	0.53	TTC1980	1323	0.58	452	97%	0.05	232	63	51.33	13.94
TSC_C10240	552	0.48	TTC0511	549	0.54	186	98%	0.05	121	34	65.05	18.28
TSC_C21330	585	0.52	TTC1236	585	0.58	195	99%	0.05	108	33	55.38	16.92
TSC_C13900	1305	0.49	TTC0674	1305	0.54	434	100%	0.05	209	50	48.16	11.52
TSC_C01860	207	0.34	TTC1549	147	0.39	68	81%	0.05	33	9	48.53	13.24
TSC_C14910	606	0.45	TTC0294	597	0.50	207	96%	0.05	140	49	67.63	23.67
TSC_C19110	741	0.53	TTC0084	741	0.58	250	98%	0.05	135	45	54.00	18.00
TSC_C23120	636	0.44	TTC1259	624	0.49	216	96%	0.05	129	45	59.72	20.83
TSC_C02610	681	0.45	TTC0648	687	0.50	228	99%	0.05	126	31	55.26	13.60
TSC_C03750	390	0.43	TTC1369	390	0.48	130	98%	0.05	90	27	69.23	20.77
TSC_C15060	1473	0.49	TTC0809	1473	0.54	512	95%	0.05	292	94	57.03	18.36
TSC_C09860	543	0.48	TTC0751	537	0.53	185	96%	0.05	118	34	63.78	18.38
TSC_C04950	654	0.44	TTC0987	648	0.49	217	99%	0.05	94	17	43.32	7.83
TSC_C23880	516	0.49	TTC1693	516	0.54	171	99%	0.05	104	23	60.82	13.45
TSC_C01530	1410	0.46	TTC1620	1410	0.51	472	99%	0.05	215	48	45.55	10.17
TSC_C24040	867	0.53	TTC1520	864	0.58	290	99%	0.05	139	34	47.93	11.72
TSC_C00940	702	0.48	TTC1670	720	0.53	244	96%	0.05	141	41	57.79	16.80
TSC_C24920	2349	0.50	TTC1595	2358	0.55	804	97%	0.05	461	130	57.34	16.17
TSC_C14540	585	0.50	TTC0722	585	0.55	212	90%	0.05	163	56	76.89	26.42

TSC_C21010	1284	0.48	TTC1046	1293	0.53	433	99%	0.05	182	41	42.03	9.47
TSC_C09250	420	0.46	TTC0676	429	0.51	142	98%	0.05	46	5	32.39	3.52
TSC_C07310	897	0.51	TTC0458	900	0.56	302	98%	0.05	165	49	54.64	16.23
TSC_C17010	1011	0.49	TTC0503	1068	0.54	356	97%	0.05	178	45	50.00	12.64
TSC_C00390	672	0.50	TTC1643	693	0.55	235	96%	0.05	128	47	54.47	20.00
TSC_C05280	1119	0.49	TTC0948	1131	0.55	376	99%	0.05	164	38	43.62	10.11
TSC_C11800	621	0.48	TTC0315	621	0.53	207	99%	0.05	106	22	51.21	10.63
TSC_C19400	1683	0.47	TTC1192	1662	0.52	560	99%	0.05	214	42	38.21	7.50
TSC_C23560	420	0.51	TTC1667	444	0.57	151	94%	0.05	114	38	75.50	25.17
TSC_C00010	1344	0.42	TTC1608	1341	0.47	450	99%	0.05	196	49	43.56	10.89
TSC_C08070	1014	0.49	TTC0172	1011	0.54	343	98%	0.05	196	57	57.14	16.62
TSC_C18940	1269	0.47	TTC1087	1269	0.52	423	100%	0.05	166	35	39.24	8.27
TSC_C23130	756	0.45	TTC1260	729	0.50	256	96%	0.05	129	41	50.39	16.02
TSC_C04030	387	0.48	TTC1478	408	0.53	137	95%	0.05	71	23	51.82	16.79
TSC_C23910	1071	0.54	TTC1690	1071	0.59	359	99%	0.05	179	53	49.86	14.76
TSC_C20050	1155	0.47	TTC1893	1134	0.52	396	96%	0.05	277	75	69.95	18.94
TSC_C19020	1152	0.50	TTC1094	1146	0.55	385	99%	0.05	207	70	53.77	18.18
TSC_C17540	654	0.51	TTC1072	651	0.56	251	84%	0.05	212	96	84.46	38.25
TSC_C15480	1317	0.47	TTC0356	1317	0.52	442	99%	0.05	164	36	37.10	8.14
TSC_C19740	1008	0.45	TTC0096	1008	0.50	339	99%	0.05	141	30	41.59	8.85
TSC_C00340	468	0.47	TTC1637	450	0.52	155	97%	0.05	51	15	32.90	9.68
TSC_C04090	1080	0.48	TTC1473	1086	0.53	365	98%	0.05	178	48	48.77	13.15
TSC_C08660	1899	0.52	TTC0919	1890	0.57	645	98%	0.05	347	90	53.80	13.95
TSC_C13230	2943	0.50	TTC0533	2937	0.55	987	99%	0.05	528	123	53.50	12.46
TSC_C10640	990	0.47	TTC0421	990	0.52	399	78%	0.05	332	136	83.21	34.09
TSC_C18090	420	0.47	TTC0473	402	0.52	139	97%	0.05	47	12	33.81	8.63
TSC_C05470	1314	0.49	TTC0133	1299	0.53	441	98%	0.05	220	70	49.89	15.87
TSC_C20510	1383	0.47	TTC1001	1389	0.52	464	99%	0.05	195	53	42.03	11.42
TSC_C14550	1245	0.49	TTC0723	1236	0.54	418	98%	0.05	215	51	51.44	12.20

TSC_C00660	504	0.50	TTC1585	501	0.55	170	97%	0.05	88	23	51.76	13.53
TSC_C11010	1119	0.46	TTC0772	1149	0.51	459	78%	0.05	330	126	71.90	27.45
TSC_C01800	522	0.49	TTC1777	522	0.54	174	99%	0.05	67	17	38.51	9.77
TSC_C13610	1170	0.52	TTC0550	1173	0.57	396	98%	0.05	186	47	46.97	11.87
TSC_C19840	2859	0.48	TTC1075	2859	0.53	953	100%	0.05	340	64	35.68	6.72
TSC_C11830	963	0.52	TTC0318	939	0.57	328	96%	0.05	221	64	67.38	19.51
TSC_C13130	1263	0.49	TTC0788	1227	0.54	425	97%	0.05	214	59	50.35	13.88
TSC_C07260	681	0.51	TTC0453	687	0.56	238	95%	0.05	123	29	51.68	12.18
TSC_C01680	879	0.49	TTC1656	891	0.54	298	98%	0.05	116	29	38.93	9.73
TSC_C06290	402	0.45	TTC0143	378	0.50	138	92%	0.05	49	9	35.51	6.52
TSC_C20850	879	0.52	TTC1028	879	0.57	297	98%	0.05	167	39	56.23	13.13
TSC_C24470	1098	0.47	TTC1674	1098	0.52	373	98%	0.05	220	71	58.98	19.03
TSC_C07580	1671	0.51	TTC0871	1668	0.56	563	98%	0.05	236	54	41.92	9.59
TSC_C12120	915	0.49	TTC0610	942	0.54	319	96%	0.05	151	32	47.34	10.03
TSC_C22100	636	0.45	TTC1293	636	0.50	213	99%	0.05	99	13	46.48	6.10
TSC_C02230	813	0.49	TTC1741	792	0.54	273	97%	0.05	131	37	47.99	13.55
TSC_C01120	357	0.39	TTC0840	342	0.44	124	92%	0.05	123	49	99.19	39.52
TSC_C24890	1050	0.49	TTC1941	1059	0.54	363	96%	0.05	232	59	63.91	16.25
TSC_C09830	456	0.44	TTC0754	456	0.49	157	95%	0.05	90	27	57.32	17.20
TSC_C00040	1425	0.50	TTC1611	1425	0.55	481	98%	0.05	214	56	44.49	11.64
TSC_C17080	441	0.53	TTC0496	429	0.58	154	92%	0.05	100	38	64.94	24.68
TSC_C09960	303	0.48	TTC0269	303	0.53	101	98%	0.05	71	22	70.30	21.78
TSC_C22660	1551	0.49	TTC1213	1551	0.54	522	99%	0.05	247	75	47.32	14.37
TSC_C24620	1029	0.52	TTC1948	903	0.57	363	87%	0.05	219	67	60.33	18.46
TSC_C00130	819	0.50	TTC1795	885	0.55	295	95%	0.05	133	43	45.08	14.58
TSC_C01260	495	0.48	TTC1642	495	0.53	164	99%	0.05	81	17	49.39	10.37
TSC_C03670	969	0.55	TTC1378	969	0.60	331	97%	0.05	169	45	51.06	13.60
TSC_C13340	1110	0.49	TTC0519	1110	0.54	370	99%	0.05	155	39	41.89	10.54
TSC_C09470	1296	0.51	TTC0556	1305	0.55	444	97%	0.05	301	102	67.79	22.97

TSC_C19960	924	0.45	TTC1414	918	0.50	323	94%	0.05	230	98	71.21	30.34
TSC_C15150	1548	0.52	TTC0817	1548	0.57	525	98%	0.05	275	68	52.38	12.95
TSC_C08740	300	0.44	TTC0911	300	0.49	99	99%	0.05	39	9	39.39	9.09
TSC_C11860	759	0.53	TTC0235	762	0.57	255	99%	0.05	120	25	47.06	9.80
TSC_C10300	459	0.45	TTC0583	459	0.50	154	98%	0.05	70	17	45.45	11.04
TSC_C01850	1131	0.45	TTC1550	1131	0.50	376	100%	0.05	127	25	33.78	6.65
TSC_C08900	669	0.50	TTC0384	675	0.55	232	96%	0.05	151	39	65.09	16.81
TSC_C00240	615	0.48	TTC1613	615	0.53	208	98%	0.05	135	42	64.90	20.19
TSC_C20180	639	0.53	TTC0089	627	0.58	215	97%	0.05	129	41	60.00	19.07
TSC_C23770	747	0.48	TTC1342	747	0.53	251	98%	0.05	121	38	48.21	15.14
TSC_C01360	471	0.44	TTC1789	480	0.48	162	97%	0.05	115	35	70.99	21.60
TSC_C00750	915	0.45	TTC1592	915	0.50	306	99%	0.05	136	32	44.44	10.46
TSC_C04040	1014	0.50	TTC1477	1020	0.55	344	98%	0.05	161	46	46.80	13.37
TSC_C21270	1956	0.52	TTC1230	1959	0.56	672	97%	0.05	361	105	53.72	15.63
TSC_C00090	933	0.51	TTC1799	936	0.56	324	95%	0.05	200	52	61.73	16.05
TSC_C22540	1566	0.52	TTC1772	1569	0.57	534	97%	0.05	262	76	49.06	14.23
TSC_C18110	612	0.44	TTC0471	636	0.48	213	97%	0.05	104	27	48.83	12.68
TSC_C14120	774	0.49	TTC0244	768	0.53	265	96%	0.05	140	39	52.83	14.72
TSC_C12140	588	0.47	TTC0609	612	0.51	204	97%	0.05	110	30	53.92	14.71
TSC_C12020	1593	0.52	TTC0403	1575	0.57	536	98%	0.05	309	82	57.65	15.30
TSC_C20060	654	0.47	TTC1892	663	0.52	220	99%	0.05	122	27	55.45	12.27
TSC_C05880	183	0.33	TTC0050	183	0.37	60	98%	0.05	20	2	33.33	3.33
TSC_C18590	1056	0.49	TTC1167	1071	0.54	364	97%	0.05	169	48	46.43	13.19
TSC_C07420	249	0.39	TTC0468	258	0.44	90	91%	0.05	53	19	58.89	21.11
TSC_C05110	1659	0.48	TTC0964	1659	0.52	561	98%	0.05	227	54	40.46	9.63
TSC_C03060	306	0.44	TTC1424	306	0.49	101	99%	0.05	54	14	53.47	13.86
TSC_C07180	1368	0.53	TTC0443	1356	0.58	465	97%	0.05	271	84	58.28	18.06
TSC_C21220	2211	0.56	TTC0639	2214	0.61	777	94%	0.05	504	158	64.86	20.33
TSC_C09100	1479	0.50	TTC0342	1479	0.54	497	99%	0.05	258	64	51.91	12.88

TSC_C08010	1221	0.50	TTC0900	1221	0.55	411	99%	0.05	211	58	51.34	14.11
TSC_C13050	423	0.50	TTC0782	462	0.54	167	85%	0.05	118	43	70.66	25.75
TSC_C10610	1302	0.44	TTC0416	1314	0.49	508	83%	0.05	406	159	79.92	31.30
TSC_C03250	693	0.49	TTC1807	699	0.54	235	98%	0.05	138	39	58.72	16.60
TSC_C01420	1818	0.45	TTC1631	1827	0.50	618	98%	0.05	344	122	55.66	19.74
TSC_C13950	2586	0.53	TTC0260	2577	0.58	872	98%	0.05	498	116	57.11	13.30
TSC_C21760	1311	0.49	TTC1149	1311	0.53	441	99%	0.05	206	50	46.71	11.34
TSC_C18530	1779	0.53	TTC1171	1800	0.58	611	97%	0.05	370	112	60.56	18.33
TSC_C15330	1047	0.48	TTC0836	1038	0.52	353	98%	0.05	203	51	57.51	14.45
TSC_C18600	441	0.45	TTC1177	441	0.50	152	95%	0.05	100	28	65.79	18.42
TSC_C07570	393	0.49	TTC0872	387	0.54	136	94%	0.05	97	38	71.32	27.94
TSC_C19290	579	0.51	TTC1224	552	0.56	195	95%	0.05	104	22	53.33	11.28
TSC_C09300	363	0.43	TTC0681	363	0.48	123	97%	0.05	70	22	56.91	17.89
TSC_C06410	954	0.47	TTC1870	957	0.52	319	99%	0.05	104	19	32.60	5.96
TSC_C13140	1239	0.51	TTC0792	1239	0.56	419	98%	0.05	248	66	59.19	15.75
TSC_C03190	1275	0.52	TTC1412	1272	0.57	428	99%	0.05	197	51	46.03	11.92
TSC_C03770	243	0.37	TTC1367	240	0.42	85	92%	0.05	51	17	60.00	20.00
TSC_C15260	1746	0.50	TTC0828	1728	0.55	590	98%	0.05	311	86	52.71	14.58
TSC_C20930	561	0.47	TTC1037	537	0.51	187	97%	0.05	95	28	50.80	14.97
TSC_C20230	678	0.50	TTC0982	681	0.54	232	97%	0.05	150	44	64.66	18.97
TSC_C12160	522	0.47	TTC0607	528	0.52	181	95%	0.05	82	26	45.30	14.36
TSC_C23750	699	0.46	TTC1341	693	0.50	236	97%	0.05	143	43	60.59	18.22
TSC_C15050	2457	0.52	TTC0808	2460	0.57	836	98%	0.05	500	129	59.81	15.43
TSC_C03410	468	0.47	TTC1822	486	0.52	162	97%	0.05	76	21	46.91	12.96
TSC_C23010	717	0.50	TTC1847	732	0.54	251	95%	0.05	159	50	63.35	19.92
TSC_C22430	618	0.49	TTC1327	618	0.54	205	100%	0.04	89	19	43.41	9.27
TSC_C21940	1428	0.49	TTC1198	1251	0.53	476	93%	0.04	215	68	45.17	14.29
TSC_C20500	711	0.49	TTC1000	711	0.53	238	99%	0.04	117	37	49.16	15.55
TSC_C13990	471	0.43	TTC0257	468	0.47	156	99%	0.04	58	7	37.18	4.49

TSC_C05580	2127	0.52	TTC1062	2079	0.56	718	97%	0.04	332	92	46.24	12.81
TSC_C08750	600	0.54	TTC0910	567	0.58	207	93%	0.04	119	42	57.49	20.29
TSC_C06150	936	0.52	TTC0056	936	0.56	322	96%	0.04	175	41	54.35	12.73
TSC_C24150	1203	0.50	TTC1526	1224	0.54	411	98%	0.04	209	55	50.85	13.38
TSC_C18020	1023	0.52	TTC0738	1023	0.56	346	98%	0.04	203	59	58.67	17.05
TSC_C09270	876	0.49	TTC0678	876	0.53	299	97%	0.04	196	61	65.55	20.40
TSC_C18310	906	0.48	TTC0427	909	0.52	304	99%	0.04	135	35	44.41	11.51
TSC_C03520	1089	0.47	TTC1396	1086	0.52	362	100%	0.04	149	38	41.16	10.50
TSC_C21250	1305	0.56	TTC1228	1305	0.60	451	96%	0.04	294	88	65.19	19.51
TSC_C09280	1476	0.44	TTC0679	1479	0.48	494	99%	0.04	200	40	40.49	8.10
TSC_C09910	1140	0.51	TTC0745	1101	0.56	382	97%	0.04	230	66	60.21	17.28
TSC_C20330	1185	0.50	TTC1121	1185	0.54	404	97%	0.04	202	67	50.00	16.58
TSC_C02780	219	0.38	TTC1705	219	0.42	72	99%	0.04	34	11	47.22	15.28
TSC_C15400	1257	0.51	TTC0360	1266	0.55	422	99%	0.04	160	39	37.91	9.24
TSC_C21340	411	0.51	TTC1237	384	0.55	146	88%	0.04	94	32	64.38	21.92
TSC_C22600	537	0.40	TTC1766	546	0.45	181	99%	0.04	58	11	32.04	6.08
TSC_C24970	531	0.45	TTC1600	528	0.50	178	98%	0.04	52	4	29.21	2.25
TSC_C01250	1131	0.50	TTC1641	1131	0.54	381	98%	0.04	171	45	44.88	11.81
TSC_C00030	1269	0.49	TTC1610	1269	0.53	423	100%	0.04	178	31	42.08	7.33
TSC_C19410	864	0.49	TTC1162	861	0.53	294	97%	0.04	166	46	56.46	15.65
TSC_C08580	1239	0.53	TTC0370	1239	0.58	412	100%	0.04	200	49	48.54	11.89
TSC_C14920	2676	0.46	TTC0805	2589	0.50	899	97%	0.04	367	71	40.82	7.90
TSC_C03070	258	0.45	TTC1423	258	0.50	85	99%	0.04	31	3	36.47	3.53
TSC_C24660	1050	0.45	TTC1574	1068	0.49	356	99%	0.04	156	47	43.82	13.20
TSC_C19610	1434	0.50	TTC0121	1428	0.55	496	96%	0.04	301	89	60.69	17.94
TSC_C18660	1779	0.46	TTC1906	1779	0.50	596	99%	0.04	295	74	49.50	12.42
TSC_C17770	975	0.55	TTC0413	993	0.59	332	98%	0.04	193	50	58.13	15.06
TSC_C13170	1446	0.44	TTC0537	1443	0.49	486	99%	0.04	226	49	46.50	10.08
TSC_C05310	930	0.47	TTC0945	840	0.51	316	92%	0.04	187	64	59.18	20.25

TSC_C15170	2133	0.52	TTC0819	2088	0.57	729	96%	0.04	409	103	56.10	14.13
TSC_C02160	813	0.49	TTC1782	813	0.54	276	97%	0.04	165	54	59.78	19.57
TSC_C15410	1746	0.49	TTC0359	1743	0.53	582	100%	0.04	277	74	47.59	12.71
TSC_C07340	1914	0.53	TTC0461	1875	0.57	677	92%	0.04	470	178	69.42	26.29
TSC_C00630	450	0.52	TTC1581	444	0.57	154	95%	0.04	79	20	51.30	12.99
TSC_C06240	957	0.48	TTC0139	957	0.52	323	98%	0.04	209	56	64.71	17.34
TSC_C06570	1296	0.52	TTC1854	1299	0.56	440	98%	0.04	267	67	60.68	15.23
TSC_C08610	1155	0.53	TTC0373	1146	0.58	398	96%	0.04	246	67	61.81	16.83
TSC_C05780	927	0.53	TTC0040	903	0.57	318	95%	0.04	174	53	54.72	16.67
TSC_C00470	636	0.52	TTC1935	636	0.56	213	99%	0.04	126	27	59.15	12.68
TSC_C20600	366	0.47	TTC1009	375	0.51	124	98%	0.04	81	19	65.32	15.32
TSC_C02170	891	0.48	TTC1781	897	0.52	302	98%	0.04	171	49	56.62	16.23
TSC_C17090	687	0.52	TTC0495	687	0.56	228	100%	0.04	114	36	50.00	15.79
TSC_C15350	906	0.46	TTC0838	906	0.51	303	99%	0.04	153	32	50.50	10.56
TSC_C21630	1407	0.51	TTC1139	1407	0.55	471	99%	0.04	249	70	52.87	14.86
TSC_C09220	2076	0.57	TTC0696	2121	0.61	725	96%	0.04	486	143	67.03	19.72
TSC_C09450	1116	0.51	TTC0555	1113	0.56	374	99%	0.04	191	49	51.07	13.10
TSC_C21550	795	0.50	TTC1928	795	0.54	269	98%	0.04	139	39	51.67	14.50
TSC_C06550	1818	0.54	TTC1856	1821	0.58	611	99%	0.04	331	99	54.17	16.20
TSC_C09630	1731	0.46	TTC0775	1722	0.50	578	99%	0.04	176	41	30.45	7.09
TSC_C09440	2241	0.50	TTC0554	2235	0.54	755	99%	0.04	377	86	49.93	11.39
TSC_C00540	798	0.50	TTC1578	795	0.55	270	98%	0.04	152	30	56.30	11.11
TSC_C21890	612	0.48	TTC1193	612	0.52	207	98%	0.04	101	25	48.79	12.08
TSC_C09430	1251	0.49	TTC0553	1245	0.54	416	100%	0.04	166	40	39.90	9.62
TSC_C24020	693	0.54	TTC1518	693	0.58	233	98%	0.04	127	31	54.51	13.30
TSC_C18670	1050	0.55	TTC1905	1047	0.59	354	98%	0.04	206	59	58.19	16.67
TSC_C14310	1701	0.49	TTC0209	1701	0.53	575	98%	0.04	379	122	65.91	21.22
TSC_C09000	834	0.53	TTC0332	831	0.57	286	96%	0.04	159	52	55.59	18.18
TSC_C22920	615	0.41	TTC0632	594	0.45	215	92%	0.04	189	66	87.91	30.70

TSC_C02440	1047	0.51	TTC1762	1047	0.55	349	99%	0.04	181	55	51.86	15.76
TSC_C20900	555	0.48	TTC1034	555	0.52	185	99%	0.04	88	21	47.57	11.35
TSC_C05140	1245	0.43	TTC0961	1215	0.48	455	89%	0.04	314	139	69.01	30.55
TSC_C09980	384	0.46	TTC0267	384	0.50	128	98%	0.04	88	19	68.75	14.84
TSC_C17000	1104	0.50	TTC0504	1104	0.55	369	99%	0.04	166	40	44.99	10.84
TSC_C03850	1878	0.53	TTC1358	2397	0.57	807	86%	0.04	344	93	42.63	11.52
TSC_C00640	597	0.46	TTC1583	588	0.50	199	98%	0.04	88	27	44.22	13.57
TSC_C09730	873	0.48	TTC0764	882	0.52	294	99%	0.04	166	34	56.46	11.56
TSC_C14570	501	0.52	TTC0725	501	0.56	167	99%	0.04	76	18	45.51	10.78
TSC_C08230	735	0.49	TTC0178	708	0.53	247	96%	0.04	149	41	60.32	16.60
TSC_C20890	1305	0.47	TTC1033	1299	0.51	436	99%	0.04	193	40	44.27	9.17
TSC_C14980	840	0.51	TTC0797	828	0.55	285	97%	0.04	181	51	63.51	17.89
TSC_C06430	624	0.47	TTC1866	621	0.51	208	99%	0.04	122	29	58.65	13.94
TSC_C24680	723	0.51	TTC1826	702	0.55	250	94%	0.04	152	49	60.80	19.60
TSC_C02560	660	0.53	TTC0643	723	0.57	242	94%	0.04	107	32	44.21	13.22
TSC_C22590	711	0.51	TTC1767	708	0.55	243	96%	0.04	138	35	56.79	14.40
TSC_C07300	795	0.46	TTC0457	795	0.50	265	99%	0.04	136	30	51.32	11.32
TSC_C24380	972	0.45	TTC1657	1059	0.49	355	95%	0.04	153	40	43.10	11.27
TSC_C14530	798	0.48	TTC0721	798	0.52	267	99%	0.04	125	25	46.82	9.36
TSC_C18950	555	0.40	TTC1088	561	0.44	186	99%	0.04	43	6	23.12	3.23
TSC_C01990	1035	0.52	TTC1535	978	0.56	360	92%	0.04	213	75	59.17	20.83
TSC_C10250	387	0.41	TTC0512	387	0.45	129	98%	0.04	56	17	43.41	13.18
TSC_C05430	1404	0.50	TTC0129	1383	0.55	469	99%	0.04	233	56	49.68	11.94
TSC_C10540	423	0.47	TTC0603	426	0.51	142	98%	0.04	61	16	42.96	11.27
TSC_C08380	966	0.51	TTC0853	981	0.55	330	98%	0.04	182	50	55.15	15.15
TSC_C23060	477	0.49	TTC0014	477	0.53	161	97%	0.04	96	23	59.63	14.29
TSC_C16970	705	0.44	TTC0507	702	0.48	234	99%	0.04	58	5	24.79	2.14
TSC_C04600	192	0.36	TTC1985	192	0.40	63	98%	0.04	27	9	42.86	14.29
TSC_C24140	804	0.46	TTC1525	786	0.50	270	97%	0.04	124	28	45.93	10.37

TSC_C04160	1731	0.53	TTC1465	1725	0.57	584	98%	0.04	254	59	43.49	10.10
TSC_C20880	378	0.48	TTC1032	378	0.52	127	98%	0.04	68	18	53.54	14.17
TSC_C23170	996	0.45	TTC1264	993	0.50	337	98%	0.04	173	52	51.34	15.43
TSC_C01670	591	0.44	TTC1655	591	0.48	196	99%	0.04	87	18	44.39	9.18
TSC_C06790	1530	0.53	TTC0636	1530	0.57	530	96%	0.04	332	111	62.64	20.94
TSC_C11180	1218	0.49	TTC0956	1215	0.53	438	91%	0.04	347	125	79.22	28.54
TSC_C03390	714	0.54	TTC1820	750	0.58	250	97%	0.04	140	30	56.00	12.00
TSC_C05260	1092	0.47	TTC0950	1113	0.51	382	96%	0.04	233	71	60.99	18.59
TSC_C10010	1251	0.49	TTC0264	1251	0.53	418	99%	0.04	168	21	40.19	5.02
TSC_C05010	1053	0.51	TTC0974	1044	0.55	361	96%	0.04	251	91	69.53	25.21
TSC_C05550	1647	0.45	TTC1065	1647	0.49	551	99%	0.04	237	68	43.01	12.34
TSC_C06070	750	0.47	TTC1922	744	0.51	256	96%	0.04	148	44	57.81	17.19
TSC_C07590	735	0.50	TTC0870	735	0.54	246	99%	0.04	144	42	58.54	17.07
TSC_C18510	1113	0.50	TTC1173	750	0.54	376	78%	0.04	135	36	35.90	9.57
TSC_C05650	729	0.46	TTC0071	729	0.50	243	99%	0.04	104	25	42.80	10.29
TSC_C23260	933	0.44	TTC1274	933	0.48	310	100%	0.04	106	19	34.19	6.13
TSC_C17030	573	0.52	TTC0501	534	0.56	200	91%	0.04	161	60	80.50	30.00
TSC_C01920	897	0.50	TTC1543	846	0.54	301	96%	0.04	119	34	39.53	11.30
TSC_C22150	477	0.45	TTC1298	480	0.49	168	93%	0.04	86	26	51.19	15.48
TSC_C05600	420	0.38	TTC0067	426	0.42	147	94%	0.04	78	22	53.06	14.97
TSC_C16960	591	0.43	TTC0508	591	0.47	196	99%	0.04	53	11	27.04	5.61
TSC_C22170	942	0.47	TTC1300	948	0.51	322	97%	0.04	136	30	42.24	9.32
TSC_C07620	1038	0.50	TTC0867	1038	0.53	345	100%	0.04	156	34	45.22	9.86
TSC_C13450	1143	0.48	TTC0407	1143	0.52	383	99%	0.04	159	37	41.51	9.66
TSC_C00290	858	0.54	TTC1527	879	0.58	341	82%	0.04	234	93	68.62	27.27
TSC_C23250	1116	0.49	TTC1273	1140	0.53	384	97%	0.04	221	74	57.55	19.27
TSC_C13390	1593	0.48	TTC0513	1593	0.52	534	99%	0.04	237	60	44.38	11.24
TSC_C23110	693	0.50	TTC1258	693	0.54	231	99%	0.04	120	27	51.95	11.69
TSC_C22990	1128	0.51	TTC1848	1179	0.55	402	95%	0.04	219	64	54.48	15.92

TSC_C21690	753	0.52	TTC1141	747	0.56	260	95%	0.04	172	52	66.15	20.00
TSC_C09420	399	0.48	TTC0552	399	0.52	132	99%	0.04	71	14	53.79	10.61
TSC_C05320	1074	0.50	TTC0944	1074	0.54	365	97%	0.04	225	53	61.64	14.52
TSC_C14950	315	0.46	TTC0802	318	0.49	112	92%	0.04	88	30	78.57	26.79
TSC_C06360	882	0.41	TTC1875	888	0.45	295	99%	0.04	116	22	39.32	7.46
TSC_C18360	2388	0.47	TTC0418	2388	0.51	796	100%	0.04	292	52	36.68	6.53
TSC_C04150	1023	0.48	TTC1466	1023	0.52	341	99%	0.04	129	25	37.83	7.33
TSC_C03660	1005	0.49	TTC1379	1005	0.53	338	99%	0.04	230	69	68.05	20.41
TSC_C23660	630	0.49	TTC1568	633	0.53	212	98%	0.04	98	29	46.23	13.68
TSC_C02220	306	0.43	TTC1740	306	0.47	101	99%	0.04	32	16	31.68	15.84
TSC_C15080	1305	0.49	TTC0811	1305	0.53	434	100%	0.04	221	50	50.92	11.52
TSC_C25030	813	0.54	TTC1604	810	0.58	271	99%	0.04	160	49	59.04	18.08
TSC_C23380	1329	0.42	TTC1287	1329	0.46	452	98%	0.04	228	66	50.44	14.60
TSC_C02490	2331	0.56	TTC1354	2217	0.60	823	91%	0.04	531	197	64.52	23.94
TSC_C00570	414	0.49	TTC1580	411	0.53	149	90%	0.04	117	41	78.52	27.52
TSC_C06800	498	0.53	TTC0635	498	0.57	168	98%	0.04	126	39	75.00	23.21
TSC_C24260	1038	0.47	TTC1555	978	0.51	349	96%	0.04	130	27	37.25	7.74
TSC_C15040	765	0.43	TTC0807	765	0.47	257	98%	0.04	125	38	48.64	14.79
TSC_C09520	1854	0.49	TTC0217	1854	0.52	620	99%	0.04	273	54	44.03	8.71
TSC_C10020	1068	0.49	TTC0263	1068	0.52	361	98%	0.04	230	65	63.71	18.01
TSC_C14060	1215	0.50	TTC0249	1215	0.54	412	98%	0.04	204	68	49.51	16.50
TSC_C01320	429	0.54	TTC1444	417	0.58	144	96%	0.04	89	30	61.81	20.83
TSC_C18990	366	0.37	TTC1092	366	0.41	121	99%	0.04	52	13	42.98	10.74
TSC_C17150	792	0.48	TTC0487	792	0.52	269	97%	0.04	185	53	68.77	19.70
TSC_C07800	636	0.50	TTC0195	636	0.53	212	99%	0.04	88	28	41.51	13.21
TSC_C14940	2019	0.53	TTC0803	2001	0.57	690	97%	0.04	438	126	63.48	18.26
TSC_C23080	843	0.46	TTC0016	852	0.50	284	99%	0.04	139	41	48.94	14.44
TSC_C05150	2427	0.50	TTC0960	2436	0.54	816	99%	0.04	424	86	51.96	10.54
TSC_C00190	321	0.46	TTC1435	330	0.50	114	93%	0.04	75	27	65.79	23.68

TSC_C20290	837	0.49	TTC1125	843	0.52	282	99%	0.04	127	30	45.04	10.64
TSC_C03980	276	0.40	TTC1481	285	0.44	95	96%	0.04	63	22	66.32	23.16
TSC_C04060	492	0.45	TTC1475	498	0.49	168	97%	0.04	93	27	55.36	16.07
TSC_C02550	789	0.59	TTC0642	756	0.63	268	95%	0.04	158	53	58.96	19.78
TSC_C21380	1809	0.52	TTC1241	1857	0.55	628	97%	0.04	321	87	51.11	13.85
TSC_C04870	648	0.40	TTC0995	648	0.44	216	99%	0.04	103	29	47.69	13.43
TSC_C11750	741	0.53	TTC0309	741	0.56	253	97%	0.04	132	37	52.17	14.62
TSC_C23440	402	0.47	TTC1494	402	0.50	136	97%	0.04	124	43	91.18	31.62
TSC_C13320	279	0.43	TTC0521	279	0.46	94	97%	0.04	52	15	55.32	15.96
TSC_C01330	2010	0.47	TTC1786	2013	0.50	675	99%	0.04	352	96	52.15	14.22
TSC_C23740	1356	0.52	TTC1339	1356	0.56	453	99%	0.04	182	45	40.18	9.93
TSC_C17830	1293	0.51	TTC0705	1236	0.55	445	94%	0.04	317	91	71.24	20.45
TSC_C07890	399	0.47	TTC0926	399	0.50	134	98%	0.04	74	20	55.22	14.93
TSC_C21090	783	0.46	TTC1054	780	0.50	266	97%	0.04	125	28	46.99	10.53
TSC_C00250	1848	0.47	TTC1614	1848	0.51	618	99%	0.04	232	57	37.54	9.22
TSC_C09760	903	0.46	TTC0761	903	0.50	303	99%	0.04	175	49	57.76	16.17
TSC_C12200	390	0.49	TTC0617	342	0.53	131	91%	0.04	59	18	45.04	13.74
TSC_C08260	1650	0.44	TTC0181	1647	0.48	551	99%	0.04	226	51	41.02	9.26
TSC_C21110	453	0.44	TTC1056	459	0.48	152	99%	0.04	79	26	51.97	17.11
TSC_C00690	1479	0.52	TTC1588	1590	0.56	576	87%	0.04	281	87	48.78	15.10
TSC_C13980	696	0.52	TTC0258	876	0.56	303	84%	0.04	164	44	54.13	14.52
TSC_C18920	1302	0.49	TTC1085	1299	0.52	443	97%	0.04	199	49	44.92	11.06
TSC_C04320	240	0.45	TTC1451	243	0.48	82	96%	0.04	45	13	54.88	15.85
TSC_C05170	858	0.46	TTC0957	858	0.50	285	100%	0.04	99	28	34.74	9.82
TSC_C24630	525	0.51	TTC1947	672	0.55	223	87%	0.04	62	15	27.80	6.73
TSC_C17970	1125	0.46	TTC0733	1188	0.50	405	94%	0.04	170	64	41.98	15.80
TSC_C17200	762	0.49	TTC0483	753	0.53	257	97%	0.04	123	25	47.86	9.73
TSC_C08040	1503	0.51	TTC0897	1503	0.55	509	98%	0.04	250	49	49.12	9.63
TSC_C15550	882	0.48	TTC0352	882	0.52	293	100%	0.04	79	10	26.96	3.41

TSC_C21600	2397	0.49	TTC1136	2394	0.53	803	99%	0.04	338	92	42.09	11.46
TSC_C04280	1035	0.47	TTC1456	1035	0.50	348	99%	0.04	200	52	57.47	14.94
TSC_C08730	1959	0.50	TTC0912	1959	0.54	655	99%	0.04	309	86	47.18	13.13
TSC_C01230	396	0.50	TTC1639	396	0.54	135	96%	0.04	51	17	37.78	12.59
TSC_C20710	1152	0.49	TTC1018	1152	0.52	383	100%	0.04	168	31	43.86	8.09
TSC_C09310	399	0.45	TTC0682	399	0.48	135	97%	0.04	82	22	60.74	16.30
TSC_C19100	807	0.47	TTC0085	807	0.51	275	97%	0.04	134	40	48.73	14.55
TSC_C13870	516	0.52	TTC0670	489	0.56	172	96%	0.04	88	22	51.16	12.79
TSC_C20920	432	0.47	TTC1036	420	0.51	147	95%	0.04	60	8	40.82	5.44
TSC_C17930	1218	0.48	TTC0730	1257	0.52	419	98%	0.04	150	25	35.80	5.97
TSC_C21520	1314	0.50	TTC1254	1311	0.54	441	99%	0.04	203	39	46.03	8.84
TSC_C17980	1509	0.51	TTC0734	1491	0.54	508	98%	0.04	249	61	49.02	12.01
TSC_C09940	1224	0.43	TTC0271	1224	0.47	409	99%	0.04	162	31	39.61	7.58
TSC_C24940	969	0.52	TTC1597	969	0.55	324	99%	0.04	130	27	40.12	8.33
TSC_C04020	684	0.46	TTC1479	684	0.49	231	98%	0.04	139	44	60.17	19.05
TSC_C08140	996	0.48	TTC0177	996	0.52	334	99%	0.04	147	34	44.01	10.18
TSC_C04120	780	0.50	TTC1469	786	0.54	266	97%	0.04	191	50	71.80	18.80
TSC_C04270	849	0.52	TTC1457	861	0.56	291	97%	0.04	189	57	64.95	19.59
TSC_C18810	1194	0.43	TTC0009	1266	0.46	433	94%	0.04	202	43	46.65	9.93
TSC_C07160	453	0.42	TTC0441	453	0.46	151	99%	0.04	60	21	39.74	13.91
TSC_C03640	783	0.51	TTC1381	774	0.55	265	97%	0.04	160	38	60.38	14.34
TSC_C01350	618	0.49	TTC1788	612	0.52	210	97%	0.03	126	33	60.00	15.71
TSC_C07770	306	0.45	TTC1285	288	0.48	116	81%	0.03	81	31	69.83	26.72
TSC_C09690	507	0.40	TTC0769	507	0.43	168	99%	0.03	60	16	35.71	9.52
TSC_C20260	1884	0.51	TTC1128	1875	0.54	631	99%	0.03	253	51	40.10	8.08
TSC_C15220	1161	0.49	TTC0824	1131	0.53	391	97%	0.03	259	89	66.24	22.76
TSC_C07830	267	0.50	TTC0199	267	0.53	89	98%	0.03	32	8	35.96	8.99
TSC_C04110	1041	0.49	TTC1470	1050	0.53	361	96%	0.03	237	75	65.65	20.78
TSC_C01440	1152	0.50	TTC1632	1137	0.54	389	98%	0.03	215	52	55.27	13.37

TSC_C05160	1704	0.54	TTC0959	1638	0.58	586	94%	0.03	366	109	62.46	18.60
TSC_C07520	1080	0.47	TTC0877	1080	0.50	370	97%	0.03	207	56	55.95	15.14
TSC_C19500	1587	0.45	TTC0107	1587	0.48	531	99%	0.03	275	72	51.79	13.56
TSC_C16980	558	0.39	TTC0506	558	0.42	186	99%	0.03	60	16	32.26	8.60
TSC_C07740	1191	0.51	TTC0191	1191	0.55	397	99%	0.03	199	44	50.13	11.08
TSC_C13890	255	0.42	TTC0673	267	0.45	88	97%	0.03	35	6	39.77	6.82
TSC_C22070	945	0.50	TTC1291	939	0.53	319	98%	0.03	178	37	55.80	11.60
TSC_C24750	564	0.47	TTC1560	564	0.50	188	99%	0.03	69	15	36.70	7.98
TSC_C08270	762	0.50	TTC0182	762	0.54	256	98%	0.03	124	28	48.44	10.94
TSC_C22760	393	0.48	TTC0033	393	0.51	131	98%	0.03	51	8	38.93	6.11
TSC_C13910	1350	0.49	TTC0695	1353	0.52	459	98%	0.03	237	70	51.63	15.25
TSC_C18190	426	0.53	TTC0890	426	0.57	144	97%	0.03	71	19	49.31	13.19
TSC_C01220	906	0.49	TTC1638	906	0.53	305	98%	0.03	138	34	45.25	11.15
TSC_C07820	1659	0.53	TTC0198	1653	0.56	561	98%	0.03	288	84	51.34	14.97
TSC_C00320	639	0.47	TTC0652	723	0.50	244	92%	0.03	152	47	62.30	19.26
TSC_C15560	1194	0.51	TTC0351	1167	0.54	416	94%	0.03	235	64	56.49	15.38
TSC_C06370	906	0.48	TTC1874	906	0.51	301	100%	0.03	109	23	36.21	7.64
TSC_C08330	678	0.51	TTC0187	678	0.55	241	92%	0.03	149	48	61.83	19.92
TSC_C12650	1434	0.53	TTC0634	1428	0.56	481	99%	0.03	183	39	38.05	8.11
TSC_C17240	954	0.47	TTC0231	954	0.50	317	100%	0.03	142	40	44.79	12.62
TSC_C24420	249	0.46	TTC1672	243	0.50	82	98%	0.03	62	22	75.61	26.83
TSC_C22740	1908	0.47	TTC1222	1923	0.50	644	99%	0.03	215	38	33.39	5.90
TSC_C20800	1299	0.46	TTC1025	1296	0.49	434	99%	0.03	187	44	43.09	10.14
TSC_C05270	1197	0.48	TTC0949	1200	0.51	409	97%	0.03	231	55	56.48	13.45
TSC_C04340	1377	0.49	TTC1966	1350	0.52	460	98%	0.03	201	46	43.70	10.00
TSC_C00560	591	0.43	TTC1579	585	0.46	197	98%	0.03	108	42	54.82	21.32
TSC_C22050	1275	0.49	TTC1211	1275	0.52	425	100%	0.03	173	31	40.71	7.29
TSC_C02820	939	0.52	TTC1708	936	0.55	316	98%	0.03	166	50	52.53	15.82
TSC_C24490	297	0.38	TTC1964	297	0.41	99	98%	0.03	31	6	31.31	6.06

TSC_C04310	360	0.47	TTC1452	378	0.51	126	96%	0.03	43	6	34.13	4.76
TSC_C18230	1395	0.49	TTC0433	1395	0.53	470	98%	0.03	202	63	42.98	13.40
TSC_C22310	186	0.43	TTC1315	186	0.46	61	98%	0.03	20	2	32.79	3.28
TSC_C04050	837	0.44	TTC1476	846	0.47	283	98%	0.03	134	32	47.35	11.31
TSC_C09770	555	0.45	TTC0760	555	0.48	185	99%	0.03	69	20	37.30	10.81
TSC_C02420	429	0.50	TTC1760	426	0.53	145	97%	0.03	60	12	41.38	8.28
TSC_C01520	1002	0.52	TTC1619	1002	0.56	337	99%	0.03	189	51	56.08	15.13
TSC_C23730	399	0.49	TTC1338	399	0.52	132	99%	0.03	59	17	44.70	12.88
TSC_C02350	1386	0.51	TTC1753	1386	0.55	465	99%	0.03	219	61	47.10	13.12
TSC_C13330	1269	0.49	TTC0520	1266	0.53	427	98%	0.03	217	62	50.82	14.52
TSC_C20420	633	0.44	TTC1110	633	0.48	223	93%	0.03	129	43	57.85	19.28
TSC_C24710	1272	0.50	TTC1564	1242	0.53	425	98%	0.03	202	51	47.53	12.00
TSC_C14810	1236	0.47	TTC0328	1245	0.50	416	99%	0.03	141	29	33.89	6.97
TSC_C24560	855	0.53	TTC1955	855	0.56	289	98%	0.03	150	42	51.90	14.53
TSC_C15010	1974	0.50	TTC0794	1974	0.53	663	99%	0.03	343	73	51.73	11.01
TSC_C03510	1617	0.52	TTC1397	1596	0.56	561	95%	0.03	333	112	59.36	19.96
TSC_C17900	486	0.46	TTC0726	486	0.49	161	99%	0.03	65	8	40.37	4.97
TSC_C23520	810	0.51	TTC1509	807	0.54	276	97%	0.03	182	53	65.94	19.20
TSC_C04190	1032	0.47	TTC1464	1038	0.50	345	99%	0.03	131	23	37.97	6.67
TSC_C07730	471	0.44	TTC0854	471	0.47	157	99%	0.03	116	37	73.89	23.57
TSC_C19090	465	0.53	TTC0086	450	0.56	157	96%	0.03	70	18	44.59	11.46
TSC_C14560	1056	0.48	TTC0724	1059	0.51	353	99%	0.03	134	23	37.96	6.52
TSC_C18100	954	0.45	TTC0472	945	0.48	320	98%	0.03	126	34	39.38	10.63
TSC_C24500	450	0.44	TTC1963	453	0.47	154	96%	0.03	79	25	51.30	16.23
TSC_C02800	1632	0.50	TTC1707	1632	0.53	545	99%	0.03	191	44	35.05	8.07
TSC_C07140	1221	0.45	TTC0440	1326	0.48	441	96%	0.03	164	34	37.19	7.71
TSC_C01140	885	0.47	TTC1728	891	0.51	298	99%	0.03	116	33	38.93	11.07
TSC_C24350	714	0.50	TTC1660	726	0.53	243	98%	0.03	136	34	55.97	13.99
TSC_C10420	891	0.48	TTC0591	900	0.51	301	98%	0.03	118	26	39.20	8.64

TSC_C02450	804	0.51	TTC1763	792	0.54	278	95%	0.03	174	60	62.59	21.58
TSC_C08800	672	0.50	TTC0905	672	0.53	223	100%	0.03	84	18	37.67	8.07
TSC_C21310	981	0.48	TTC1234	852	0.52	326	93%	0.03	144	33	44.17	10.12
TSC_C22620	1227	0.49	TTC1764	1227	0.52	411	99%	0.03	173	38	42.09	9.25
TSC_C00820	306	0.40	TTC1713	306	0.43	101	99%	0.03	20	4	19.80	3.96
TSC_C18220	900	0.51	TTC0434	879	0.54	301	98%	0.03	172	55	57.14	18.27
TSC_C21300	1206	0.48	TTC1233	1194	0.51	404	99%	0.03	222	52	54.95	12.87
TSC_C23390	1320	0.41	TTC1286	1320	0.44	440	100%	0.03	172	41	39.09	9.32
TSC_C10330	1566	0.53	TTC0586	1566	0.56	522	100%	0.03	216	44	41.38	8.43
TSC_C14340	744	0.52	TTC0205	744	0.55	253	97%	0.03	150	35	59.29	13.83
TSC_C23590	1980	0.47	TTC1516	1980	0.50	662	99%	0.03	270	69	40.79	10.42
TSC_C21920	300	0.40	TTC1196	300	0.43	99	99%	0.03	37	9	37.37	9.09
TSC_C00910	1824	0.48	TTC1673	1827	0.51	620	98%	0.03	281	64	45.32	10.32
TSC_C19680	984	0.46	TTC0102	984	0.49	329	99%	0.03	120	19	36.47	5.78
TSC_C12710	1203	0.48	TTC0623	1206	0.51	402	99%	0.03	178	34	44.28	8.46
TSC_C01600	384	0.50	TTC1648	384	0.53	130	97%	0.03	72	24	55.38	18.46
TSC_C05860	918	0.51	TTC0048	918	0.54	308	99%	0.03	141	32	45.78	10.39
TSC_C03300	1050	0.55	TTC1812	1053	0.58	355	98%	0.03	198	60	55.77	16.90
TSC_C18420	474	0.43	TTC1683	489	0.46	192	79%	0.03	134	59	69.79	30.73
TSC_C06460	546	0.40	TTC1864	546	0.43	181	99%	0.03	71	15	39.23	8.29
TSC_C00150	1209	0.49	TTC1793	1209	0.52	409	98%	0.03	212	46	51.83	11.25
TSC_C07660	504	0.47	TTC0859	513	0.50	180	93%	0.03	137	50	76.11	27.78
TSC_C06080	645	0.46	TTC1923	639	0.49	214	99%	0.03	103	20	48.13	9.35
TSC_C23680	1620	0.48	TTC1824	1611	0.51	546	98%	0.03	245	70	44.87	12.82
TSC_C13820	858	0.43	TTC0666	858	0.46	286	99%	0.03	107	25	37.41	8.74
TSC_C14900	1053	0.52	TTC0295	1053	0.55	352	99%	0.03	156	42	44.32	11.93
TSC_C00670	747	0.54	TTC1586	747	0.57	249	99%	0.03	153	38	61.45	15.26
TSC_C06960	522	0.47	TTC0156	528	0.50	176	98%	0.03	89	30	50.57	17.05
TSC_C17960	2025	0.49	TTC0732	2031	0.52	679	99%	0.03	309	61	45.51	8.98

TSC_C14300	639	0.44	TTC0210	636	0.47	217	97%	0.03	119	36	54.84	16.59
TSC_C21060	543	0.44	TTC1051	561	0.47	189	96%	0.03	97	30	51.32	15.87
TSC_C19700	1359	0.44	TTC0100	1347	0.47	465	96%	0.03	289	74	62.15	15.91
TSC_C05920	1281	0.48	TTC1907	1284	0.51	438	97%	0.03	229	61	52.28	13.93
TSC_C22390	342	0.44	TTC1323	342	0.47	113	99%	0.03	30	5	26.55	4.42
TSC_C20540	756	0.51	TTC1004	756	0.54	261	96%	0.03	185	59	70.88	22.61
TSC_C22400	282	0.39	TTC1324	282	0.42	93	99%	0.03	19	0	20.43	0.00
TSC_C22220	768	0.49	TTC1306	768	0.52	255	100%	0.03	98	22	38.43	8.63
TSC_C09670	480	0.42	TTC0771	471	0.45	167	93%	0.03	112	37	67.07	22.16
TSC_C21810	684	0.48	TTC1153	684	0.51	227	100%	0.03	110	29	48.46	12.78
TSC_C23490	924	0.46	TTC1505	924	0.49	310	99%	0.03	130	29	41.94	9.35
TSC_C08410	1014	0.48	TTC0850	1014	0.51	338	99%	0.03	122	29	36.09	8.58
TSC_C05890	522	0.48	TTC0051	537	0.51	180	97%	0.03	75	22	41.67	12.22
TSC_C07030	1485	0.46	TTC0164	1272	0.49	499	91%	0.03	162	39	32.46	7.82
TSC_C23050	1059	0.55	TTC1082	1038	0.58	354	98%	0.03	228	69	64.41	19.49
TSC_C03970	1143	0.53	TTC1482	1143	0.55	391	97%	0.03	226	68	57.80	17.39
TSC_C02830	228	0.43	TTC1425	354	0.46	117	77%	0.03	40	10	34.19	8.55
TSC_C03200	969	0.54	TTC1411	900	0.57	324	95%	0.03	187	59	57.72	18.21
TSC_C20160	420	0.48	TTC0091	411	0.51	140	97%	0.03	82	27	58.57	19.29
TSC_C17780	1062	0.44	TTC0412	1062	0.47	353	100%	0.03	152	34	43.06	9.63
TSC_C10220	1122	0.47	TTC0934	1140	0.50	380	99%	0.03	160	41	42.11	10.79
TSC_C03280	543	0.46	TTC1810	474	0.49	186	89%	0.03	99	37	53.23	19.89
TSC_C21930	663	0.48	TTC1197	663	0.51	222	99%	0.03	102	22	45.95	9.91
TSC_C10410	1377	0.54	TTC0590	1497	0.57	508	94%	0.03	307	90	60.43	17.72
TSC_C06260	531	0.45	TTC0141	537	0.48	194	90%	0.03	127	46	65.46	23.71
TSC_C03150	318	0.39	TTC1417	279	0.42	107	90%	0.03	57	22	53.27	20.56
TSC_C18980	396	0.35	TTC1091	396	0.38	131	99%	0.03	54	16	41.22	12.21
TSC_C03140	831	0.53	TTC1416	831	0.55	282	97%	0.03	133	28	47.16	9.93
TSC_C12190	2010	0.52	TTC0604	2007	0.55	676	99%	0.03	394	101	58.28	14.94

TSC_C07360	2502	0.52	TTC0463	2502	0.55	891	93%	0.03	535	184	60.04	20.65
TSC_C09680	1689	0.43	TTC0770	1689	0.46	562	100%	0.03	233	46	41.46	8.19
TSC_C13080	1728	0.48	TTC0785	1728	0.51	582	99%	0.03	336	79	57.73	13.57
TSC_C15070	1338	0.49	TTC0810	1341	0.52	455	98%	0.03	256	79	56.26	17.36
TSC_C21080	579	0.40	TTC1053	573	0.43	193	98%	0.03	99	27	51.30	13.99
TSC_C03710	417	0.48	TTC1374	417	0.51	139	99%	0.03	45	8	32.37	5.76
TSC_C17850	729	0.51	TTC0928	729	0.53	243	99%	0.03	97	22	39.92	9.05
TSC_C05590	726	0.56	TTC1061	723	0.59	271	87%	0.03	180	59	66.42	21.77
TSC_C02540	540	0.49	TTC0641	543	0.52	184	97%	0.03	95	33	51.63	17.93
TSC_C03820	1122	0.50	TTC1361	1122	0.53	382	97%	0.03	216	75	56.54	19.63
TSC_C09350	984	0.48	TTC0686	978	0.50	331	98%	0.03	107	12	32.33	3.63
TSC_C25010	309	0.45	TTC1602	228	0.48	105	80%	0.03	44	11	41.90	10.48
TSC_C17180	1437	0.50	TTC0485	1437	0.53	479	100%	0.03	180	39	37.58	8.14
TSC_C23990	1161	0.42	TTC1681	1170	0.45	394	98%	0.03	214	69	54.31	17.51
TSC_C14400	216	0.38	TTC0708	294	0.41	97	84%	0.03	27	4	27.84	4.12
TSC_C11810	195	0.45	TTC0316	195	0.47	65	97%	0.03	44	17	67.69	26.15
TSC_C04940	1230	0.52	TTC0988	1221	0.55	416	98%	0.03	226	71	54.33	17.07
TSC_C08830	3300	0.50	TTC0377	3300	0.53	1116	98%	0.03	550	143	49.28	12.81
TSC_C17130	1044	0.51	TTC0490	1056	0.54	354	98%	0.03	127	34	35.88	9.60
TSC_C22280	339	0.42	TTC1312	339	0.45	114	97%	0.03	38	3	33.33	2.63
TSC_C23510	789	0.53	TTC1508	768	0.55	265	97%	0.03	130	45	49.06	16.98
TSC_C24360	1143	0.52	TTC1659	1146	0.54	384	99%	0.03	177	51	46.09	13.28
TSC_C09870	675	0.46	TTC0750	675	0.48	228	98%	0.03	127	39	55.70	17.11
TSC_C03480	1005	0.43	TTC1400	999	0.46	338	98%	0.03	187	63	55.33	18.64
TSC_C01830	1119	0.51	TTC1552	1119	0.53	372	100%	0.03	153	34	41.13	9.14
TSC_C01810	378	0.45	TTC1776	378	0.48	126	98%	0.03	40	8	31.75	6.35
TSC_C21580	2001	0.52	TTC1134	1977	0.54	671	98%	0.03	255	64	38.00	9.54
TSC_C00180	480	0.47	TTC1436	474	0.50	160	98%	0.03	76	21	47.50	13.13
TSC_C06010	546	0.47	TTC1916	546	0.49	181	99%	0.03	55	8	30.39	4.42

TSC_C01570	486	0.39	TTC1646	405	0.42	162	90%	0.03	74	20	45.68	12.35
TSC_C22490	411	0.44	TTC1333	405	0.47	137	98%	0.03	34	6	24.82	4.38
TSC_C03110	1098	0.54	TTC1419	1098	0.56	378	96%	0.03	228	85	60.32	22.49
TSC_C08970	1164	0.53	TTC0238	1137	0.55	428	88%	0.03	365	147	85.28	34.35
TSC_C15700	606	0.43	TTC0480	606	0.46	203	99%	0.03	136	35	67.00	17.24
TSC_C14100	3108	0.49	TTC0247	3087	0.52	1049	98%	0.03	466	114	44.42	10.87
TSC_C05730	1935	0.47	TTC0079	1935	0.50	650	99%	0.03	250	46	38.46	7.08
TSC_C01240	1242	0.57	TTC1640	1245	0.59	426	97%	0.03	233	71	54.69	16.67
TSC_C03320	1062	0.57	TTC1813	1062	0.59	354	99%	0.03	151	30	42.66	8.47
TSC_C24930	456	0.53	TTC1596	465	0.56	168	89%	0.03	110	41	65.48	24.40
TSC_C10310	843	0.55	TTC0584	852	0.57	290	97%	0.03	193	59	66.55	20.34
TSC_C25050	741	0.54	TTC1606	750	0.56	262	94%	0.03	156	54	59.54	20.61
TSC_C13380	369	0.45	TTC0514	369	0.48	122	99%	0.03	48	5	39.34	4.10
TSC_C08290	198	0.39	TTC0184	189	0.41	65	96%	0.03	19	7	29.23	10.77
TSC_C09150	1716	0.47	TTC0347	1716	0.49	575	99%	0.03	210	48	36.52	8.35
TSC_C06930	762	0.46	TTC0153	762	0.49	259	97%	0.03	156	47	60.23	18.15
TSC_C06390	909	0.41	TTC1872	909	0.44	304	99%	0.03	81	12	26.64	3.95
TSC_C05440	1563	0.44	TTC0130	1563	0.47	522	99%	0.03	172	32	32.95	6.13
TSC_C23650	483	0.41	TTC1569	483	0.44	160	99%	0.03	75	20	46.88	12.50
TSC_C19310	924	0.49	TTC1184	924	0.52	308	99%	0.03	150	34	48.70	11.04
TSC_C03460	519	0.44	TTC1402	519	0.47	178	96%	0.03	84	32	47.19	17.98
TSC_C08780	1737	0.49	TTC0907	1737	0.52	578	100%	0.03	186	29	32.18	5.02
TSC_C08990	732	0.52	TTC0331	729	0.55	247	98%	0.03	125	27	50.61	10.93
TSC_C03950	1332	0.55	TTC1484	1332	0.57	445	99%	0.03	181	55	40.67	12.36
TSC_C06480	480	0.47	TTC1862	480	0.50	160	99%	0.03	54	12	33.75	7.50
TSC_C02310	870	0.52	TTC1749	870	0.54	290	99%	0.03	117	28	40.34	9.66
TSC_C09740	897	0.49	TTC0763	885	0.52	305	97%	0.03	185	54	60.66	17.70
TSC_C01820	582	0.49	TTC1553	567	0.52	194	98%	0.03	105	26	54.12	13.40
TSC_C21980	627	0.56	TTC1202	624	0.59	222	93%	0.03	123	45	55.41	20.27

TSC_C23580	960	0.54	TTC1668	966	0.57	329	97%	0.03	208	72	63.22	21.88
TSC_C22460	1221	0.40	TTC1734	1221	0.42	406	100%	0.03	76	8	18.72	1.97
TSC_C08050	867	0.48	TTC0169	867	0.50	288	100%	0.03	92	17	31.94	5.90
TSC_C19360	288	0.42	TTC1189	288	0.44	96	98%	0.03	29	7	30.21	7.29
TSC_C05410	1287	0.49	TTC0127	1257	0.52	430	98%	0.02	208	62	48.37	14.42
TSC_C00790	4941	0.50	TTC1806	6204	0.53	2093	87%	0.02	775	198	37.03	9.46
TSC_C04480	888	0.48	TTC1973	948	0.50	338	89%	0.02	228	77	67.46	22.78
TSC_C05990	2349	0.51	TTC1914	2352	0.54	793	99%	0.02	346	100	43.63	12.61
TSC_C01880	1266	0.53	TTC1547	1263	0.55	423	99%	0.02	166	34	39.24	8.04
TSC_C22630	990	0.56	TTC1491	990	0.58	334	98%	0.02	174	48	52.10	14.37
TSC_C22180	630	0.45	TTC1301	630	0.47	209	100%	0.02	70	14	33.49	6.70
TSC_C08400	513	0.45	TTC0851	513	0.47	170	99%	0.02	78	13	45.88	7.65
TSC_C08450	777	0.44	TTC0844	837	0.46	279	96%	0.02	123	25	44.09	8.96
TSC_C01290	1200	0.48	TTC1447	1191	0.50	410	97%	0.02	207	58	50.49	14.15
TSC_C18070	492	0.43	TTC0475	492	0.45	163	99%	0.02	66	21	40.49	12.88
TSC_C19140	972	0.48	TTC0038	1011	0.51	337	97%	0.02	133	24	39.47	7.12
TSC_C19530	1332	0.50	TTC0110	1431	0.52	485	94%	0.02	298	67	61.44	13.81
TSC_C11720	1335	0.53	TTC0137	1323	0.55	453	97%	0.02	264	77	58.28	17.00
TSC_C03800	759	0.49	TTC1363	753	0.52	276	90%	0.02	207	82	75.00	29.71
TSC_C09290	471	0.47	TTC0680	471	0.49	158	98%	0.02	71	15	44.94	9.49
TSC_C00120	357	0.49	TTC1796	357	0.51	118	99%	0.02	43	9	36.44	7.63
TSC_C00890	1266	0.50	TTC1710	1443	0.53	496	90%	0.02	237	70	47.78	14.11
TSC_C20170	774	0.41	TTC0090	774	0.44	268	95%	0.02	148	44	55.22	16.42
TSC_C07070	984	0.47	TTC0168	984	0.49	327	100%	0.02	135	33	41.28	10.09
TSC_C18900	996	0.36	TTC0999	1008	0.39	348	95%	0.02	329	108	94.54	31.03
TSC_C18710	1314	0.49	TTC1077	1332	0.52	449	98%	0.02	216	51	48.11	11.36
TSC_C20270	1860	0.48	TTC1127	1848	0.50	619	100%	0.02	183	41	29.56	6.62
TSC_C24010	609	0.49	TTC1517	624	0.51	208	98%	0.02	79	18	37.98	8.65
TSC_C22910	936	0.50	TTC1255	936	0.52	318	97%	0.02	181	53	56.92	16.67

TSC_C21720	366	0.47	TTC1144	366	0.49	124	97%	0.02	50	7	40.32	5.65
TSC_C23040	1323	0.49	TTC1083	1335	0.51	445	99%	0.02	165	32	37.08	7.19
TSC_C03490	1137	0.50	TTC1399	1134	0.52	378	100%	0.02	126	32	33.33	8.47
TSC_C01550	2667	0.48	TTC1622	2670	0.51	894	99%	0.02	378	102	42.28	11.41
TSC_C07780	2466	0.50	TTC0193	2469	0.53	843	97%	0.02	456	162	54.09	19.22
TSC_C23720	285	0.36	TTC1337	285	0.38	96	97%	0.02	42	14	43.75	14.58
TSC_C07410	1743	0.46	TTC0467	1674	0.48	602	94%	0.02	346	135	57.48	22.43
TSC_C05640	1419	0.48	TTC0070	1407	0.51	474	99%	0.02	191	48	40.30	10.13
TSC_C18470	1323	0.48	TTC0942	1398	0.50	466	97%	0.02	158	23	33.91	4.94
TSC_C21050	804	0.45	TTC1050	801	0.47	271	98%	0.02	117	32	43.17	11.81
TSC_C04460	195	0.42	TTC1970	210	0.44	71	92%	0.02	31	10	43.66	14.08
TSC_C10710	1485	0.44	TTC0971	1578	0.46	614	79%	0.02	437	199	71.17	32.41
TSC_C03580	333	0.39	TTC1385	333	0.41	111	98%	0.02	37	11	33.33	9.91
TSC_C13020	1230	0.44	TTC0779	1146	0.46	409	96%	0.02	130	20	31.78	4.89
TSC_C13310	897	0.51	TTC0522	744	0.54	300	90%	0.02	119	27	39.67	9.00
TSC_C04420	930	0.50	TTC0843	996	0.52	375	83%	0.02	350	134	93.33	35.73
TSC_C18580	1725	0.46	TTC1168	1713	0.49	577	99%	0.02	272	73	47.14	12.65
TSC_C24720	1122	0.52	TTC1563	1113	0.54	377	98%	0.02	231	57	61.27	15.12
TSC_C23310	1179	0.48	TTC1279	1296	0.50	431	95%	0.02	123	11	28.54	2.55
TSC_C10190	1506	0.54	TTC0938	1470	0.56	516	96%	0.02	312	88	60.47	17.05
TSC_C08980	1155	0.51	TTC0330	1155	0.53	385	99%	0.02	184	36	47.79	9.35
TSC_C05030	1011	0.42	TTC0972	1023	0.44	343	98%	0.02	166	39	48.40	11.37
TSC_C06270	696	0.46	TTC0142	693	0.48	232	99%	0.02	115	30	49.57	12.93
TSC_C00330	915	0.49	TTC1636	915	0.51	305	99%	0.02	112	11	36.72	3.61
TSC_C03540	207	0.41	TTC1394	204	0.44	71	93%	0.02	52	12	73.24	16.90
TSC_C21840	1392	0.50	TTC1156	1392	0.53	467	99%	0.02	200	50	42.83	10.71
TSC_C10520	456	0.50	TTC0601	411	0.52	154	92%	0.02	74	19	48.05	12.34
TSC_C14090	285	0.48	TTC0248	285	0.51	95	98%	0.02	60	14	63.16	14.74
TSC_C09590	567	0.42	TTC0224	567	0.44	188	99%	0.02	82	20	43.62	10.64

TSC_C06850	1731	0.49	TTC0144	1731	0.51	577	100%	0.02	245	57	42.46	9.88
TSC_C04570	1035	0.51	TTC1982	1029	0.53	350	98%	0.02	220	64	62.86	18.29
TSC_C24520	1158	0.49	TTC1960	1158	0.51	386	99%	0.02	147	33	38.08	8.55
TSC_C08320	465	0.52	TTC0186	471	0.55	158	97%	0.02	95	27	60.13	17.09
TSC_C06900	1317	0.52	TTC0150	1317	0.54	441	99%	0.02	238	62	53.97	14.06
TSC_C24070	807	0.47	TTC0795	735	0.49	302	82%	0.02	211	89	69.87	29.47
TSC_C08390	1689	0.49	TTC0852	1689	0.51	562	100%	0.02	202	40	35.94	7.12
TSC_C22870	1065	0.47	TTC0020	1068	0.50	355	100%	0.02	136	29	38.31	8.17
TSC_C02040	1080	0.50	TTC1716	1065	0.52	366	97%	0.02	225	64	61.48	17.49
TSC_C05530	876	0.46	TTC1068	873	0.48	292	99%	0.02	107	26	36.64	8.90
TSC_C24670	1173	0.52	TTC1575	1164	0.54	390	99%	0.02	157	42	40.26	10.77
TSC_C04810	1485	0.50	TTC0064	1482	0.52	496	99%	0.02	188	37	37.90	7.46
TSC_C14030	1524	0.49	TTC0252	1521	0.51	513	99%	0.02	215	49	41.91	9.55
TSC_C05290	1089	0.47	TTC0947	972	0.50	363	94%	0.02	134	20	36.91	5.51
TSC_C09410	288	0.48	TTC0551	288	0.50	96	98%	0.02	72	23	75.00	23.96
TSC_C13180	1734	0.50	TTC0536	1734	0.52	577	100%	0.02	223	34	38.65	5.89
TSC_C24050	1764	0.50	TTC0389	1833	0.52	732	78%	0.02	504	237	68.85	32.38
TSC_C19040	957	0.43	TTC1096	957	0.45	318	100%	0.02	117	33	36.79	10.38
TSC_C20870	294	0.42	TTC1031	321	0.44	106	95%	0.02	42	9	39.62	8.49
TSC_C00070	1431	0.48	TTC1426	1431	0.50	478	99%	0.02	191	37	39.96	7.74
TSC_C17870	1884	0.47	TTC0931	1857	0.50	634	98%	0.02	306	67	48.26	10.57
TSC_C07450	1947	0.48	TTC0884	1947	0.50	650	100%	0.02	224	51	34.46	7.85
TSC_C22240	1317	0.45	TTC1308	1317	0.47	438	100%	0.02	196	32	44.75	7.31
TSC_C02460	381	0.46	TTC1351	390	0.48	131	96%	0.02	74	19	56.49	14.50
TSC_C19640	966	0.51	TTC0106	966	0.53	322	99%	0.02	169	48	52.48	14.91
TSC_C02770	264	0.46	TTC1704	264	0.48	87	99%	0.02	31	7	35.63	8.05
TSC_C05960	540	0.47	TTC1911	540	0.49	183	97%	0.02	81	21	44.26	11.48
TSC_C22720	1224	0.53	TTC1220	1215	0.55	428	94%	0.02	253	76	59.11	17.76
TSC_C03900	1293	0.47	TTC1489	1296	0.49	432	99%	0.02	170	38	39.35	8.80

TSC_C01780	444	0.45	TTC1738	444	0.47	147	99%	0.02	32	3	21.77	2.04
TSC_C04240	4593	0.50	TTC1460	4542	0.52	1536	99%	0.02	470	107	30.60	6.97
TSC_C03160	1146	0.46	TTC1415	1107	0.48	381	98%	0.02	127	36	33.33	9.45
TSC_C24570	882	0.51	TTC1954	885	0.53	296	99%	0.02	167	42	56.42	14.19
TSC_C22320	549	0.43	TTC1316	549	0.45	182	99%	0.02	49	10	26.92	5.49
TSC_C20700	2265	0.45	TTC1017	2274	0.47	765	99%	0.02	402	116	52.55	15.16
TSC_C00270	663	0.45	TTC1616	663	0.47	226	97%	0.02	110	29	48.67	12.83
TSC_C14370	408	0.46	TTC0203	411	0.48	139	97%	0.02	64	14	46.04	10.07
TSC_C08680	675	0.47	TTC0917	675	0.49	224	100%	0.02	104	27	46.43	12.05
TSC_C20950	429	0.47	TTC1039	429	0.49	143	99%	0.02	62	19	43.36	13.29
TSC_C00720	282	0.37	TTC1589	282	0.39	93	99%	0.02	33	6	35.48	6.45
TSC_C21240	579	0.54	TTC1227	588	0.56	211	90%	0.02	126	45	59.72	21.33
TSC_C23820	1053	0.47	TTC1347	1053	0.49	352	99%	0.02	152	42	43.18	11.93
TSC_C06530	945	0.54	TTC1858	948	0.56	319	98%	0.02	155	36	48.59	11.29
TSC_C04010	2649	0.52	TTC1480	2649	0.54	891	99%	0.02	374	89	41.98	9.99
TSC_C13620	996	0.44	TTC0549	996	0.46	334	99%	0.02	127	24	38.02	7.19
TSC_C03400	501	0.51	TTC1821	483	0.53	172	94%	0.02	117	36	68.02	20.93
TSC_C22420	291	0.44	TTC1326	291	0.46	96	99%	0.02	29	7	30.21	7.29
TSC_C18630	405	0.49	TTC1180	414	0.51	138	97%	0.02	63	15	45.65	10.87
TSC_C02290	972	0.50	TTC1747	972	0.52	323	100%	0.02	134	37	41.49	11.46
TSC_C00810	1629	0.48	TTC1714	1632	0.49	543	100%	0.02	125	25	23.02	4.60
TSC_C03990	423	0.51	TTC1206	363	0.52	155	80%	0.02	115	41	74.19	26.45
TSC_C21000	549	0.43	TTC1045	558	0.45	186	98%	0.02	105	28	56.45	15.05
TSC_C15590	3132	0.47	TTC0702	3204	0.48	1075	98%	0.02	485	133	45.12	12.37
TSC_C22160	357	0.43	TTC1299	357	0.45	118	99%	0.02	36	6	30.51	5.08
TSC_C09880	939	0.48	TTC0749	939	0.50	317	98%	0.02	189	47	59.62	14.83
TSC_C20650	1005	0.46	TTC1012	1005	0.48	334	100%	0.02	123	22	36.83	6.59
TSC_C01890	492	0.45	TTC1546	492	0.47	163	99%	0.02	75	18	46.01	11.04
TSC_C15390	345	0.49	TTC0361	318	0.50	126	84%	0.02	77	24	61.11	19.05

TSC_C08060	1137	0.49	TTC0170	1176	0.51	391	98%	0.02	97	17	24.81	4.35
TSC_C19030	924	0.53	TTC1095	924	0.55	308	99%	0.02	159	43	51.62	13.96
TSC_C23550	201	0.29	TTC1512	201	0.31	66	99%	0.02	39	12	59.09	18.18
TSC_C19550	1149	0.53	TTC0112	1134	0.55	396	95%	0.02	239	86	60.35	21.72
TSC_C06810	972	0.52	TTC0201	1032	0.54	395	81%	0.02	333	137	84.30	34.68
TSC_C17110	549	0.48	TTC0492	549	0.50	198	91%	0.02	134	41	67.68	20.71
TSC_C04470	2313	0.50	TTC1972	2331	0.52	788	98%	0.02	430	135	54.57	17.13
TSC_C17750	282	0.39	TTC1809	240	0.41	93	91%	0.02	29	10	31.18	10.75
TSC_C07050	1218	0.49	TTC0166	1269	0.51	422	98%	0.02	123	18	29.15	4.27
TSC_C24730	219	0.43	TTC1561	222	0.45	80	88%	0.02	45	16	56.25	20.00
TSC_C20080	1737	0.52	TTC1890	1737	0.54	584	99%	0.02	326	102	55.82	17.47
TSC_C21490	429	0.48	TTC1251	429	0.49	143	99%	0.02	54	6	37.76	4.20
TSC_C21410	486	0.48	TTC1244	456	0.50	167	92%	0.02	74	18	44.31	10.78
TSC_C19780	192	0.34	TTC0093	192	0.36	63	98%	0.02	33	10	52.38	15.87
TSC_C23840	558	0.47	TTC1349	555	0.49	190	96%	0.02	122	39	64.21	20.53
TSC_C24030	381	0.51	TTC1519	381	0.53	129	97%	0.02	55	16	42.64	12.40
TSC_C19300	621	0.46	TTC1225	621	0.48	209	98%	0.02	108	20	51.67	9.57
TSC_C05810	1191	0.52	TTC0043	1194	0.54	402	98%	0.02	200	54	49.75	13.43
TSC_C09500	543	0.49	TTC0558	504	0.51	202	83%	0.02	145	48	71.78	23.76
TSC_C21180	1650	0.50	TTC1102	1653	0.51	555	99%	0.02	208	39	37.48	7.03
TSC_C01460	432	0.47	TTC1634	432	0.49	145	98%	0.02	82	27	56.55	18.62
TSC_C13190	1194	0.50	TTC0535	1194	0.52	398	99%	0.02	159	28	39.95	7.04
TSC_C08100	1521	0.45	TTC0175	1521	0.47	506	100%	0.02	160	23	31.62	4.55
TSC_C03260	540	0.56	TTC1808	540	0.57	187	95%	0.02	109	39	58.29	20.86
TSC_C21700	1179	0.54	TTC1142	1173	0.56	404	96%	0.02	196	66	48.51	16.34
TSC_C09900	2457	0.48	TTC0746	2415	0.49	821	99%	0.02	292	55	35.57	6.70
TSC_C23100	465	0.49	TTC1257	465	0.51	155	99%	0.02	77	30	49.68	19.35
TSC_C13470	1479	0.50	TTC0564	1284	0.52	497	92%	0.02	216	49	43.46	9.86
TSC_C02250	444	0.45	TTC1743	447	0.46	148	99%	0.02	32	7	21.62	4.73

TSC_C20120	978	0.51	TTC1886	975	0.52	326	99%	0.02	120	15	36.81	4.60
TSC_C21480	522	0.47	TTC1250	537	0.49	178	98%	0.02	77	11	43.26	6.18
TSC_C23230	1881	0.45	TTC1271	1869	0.46	628	99%	0.02	215	66	34.24	10.51
TSC_C13510	912	0.55	TTC0573	912	0.57	303	100%	0.02	32	7	10.56	2.31
TSC_C20090	1434	0.48	TTC1889	1434	0.50	479	99%	0.02	194	43	40.50	8.98
TSC_C20280	555	0.49	TTC1126	534	0.51	187	96%	0.02	98	34	52.41	18.18
TSC_C04130	1185	0.55	TTC1468	1185	0.56	397	99%	0.02	177	53	44.58	13.35
TSC_C07430	1878	0.50	TTC0886	1884	0.52	629	99%	0.02	238	55	37.84	8.74
TSC_C00460	810	0.52	TTC1936	882	0.53	300	93%	0.02	186	54	62.00	18.00
TSC_C13030	744	0.47	TTC0780	744	0.48	251	98%	0.02	100	24	39.84	9.56
TSC_C04100	627	0.50	TTC1471	654	0.51	222	95%	0.02	136	39	61.26	17.57
TSC_C23150	591	0.51	TTC1262	591	0.53	196	99%	0.02	30	4	15.31	2.04
TSC_C19650	222	0.45	TTC0105	198	0.46	77	87%	0.02	34	8	44.16	10.39
TSC_C05100	1341	0.42	TTC0965	1341	0.44	449	99%	0.02	146	27	32.52	6.01
TSC_C02300	372	0.39	TTC1748	360	0.41	126	95%	0.02	59	14	46.83	11.11
TSC_C04450	345	0.49	TTC0058	360	0.50	123	94%	0.02	76	26	61.79	21.14
TSC_C05830	1227	0.48	TTC0044	1227	0.50	409	100%	0.02	151	42	36.92	10.27
TSC_C17910	1128	0.46	TTC0727	1143	0.48	381	99%	0.02	153	28	40.16	7.35
TSC_C08090	2229	0.50	TTC0174	2211	0.51	742	99%	0.02	219	36	29.51	4.85
TSC_C04140	594	0.47	TTC1467	597	0.48	204	96%	0.02	121	34	59.31	16.67
TSC_C06180	585	0.51	TTC0061	585	0.53	195	99%	0.01	81	15	41.54	7.69
TSC_C19660	1125	0.50	TTC0104	1128	0.51	380	98%	0.01	163	33	42.89	8.68
TSC_C15290	279	0.48	TTC0832	273	0.50	95	95%	0.01	82	28	86.32	29.47
TSC_C17860	1338	0.49	TTC0929	1320	0.50	447	99%	0.01	180	44	40.27	9.84
TSC_C01470	348	0.42	TTC1635	348	0.44	119	96%	0.01	79	22	66.39	18.49
TSC_C04880	609	0.51	TTC0994	609	0.52	209	96%	0.01	102	30	48.80	14.35
TSC_C05680	1296	0.45	TTC0075	1293	0.46	434	99%	0.01	203	71	46.77	16.36
TSC_C07650	1413	0.46	TTC0865	1419	0.48	475	99%	0.01	155	27	32.63	5.68
TSC_C07020	597	0.53	TTC0162	507	0.54	200	90%	0.01	113	25	56.50	12.50

TSC_C23090	1068	0.51	TTC1256	1065	0.53	361	98%	0.01	184	51	50.97	14.13
TSC_C18140	2997	0.47	TTC0887	2994	0.49	1003	99%	0.01	320	61	31.90	6.08
TSC_C08360	1398	0.46	TTC0190	1401	0.48	469	99%	0.01	184	45	39.23	9.59
TSC_C21860	1224	0.48	TTC1160	1272	0.49	423	98%	0.01	141	23	33.33	5.44
TSC_C23140	429	0.50	TTC1261	429	0.52	144	98%	0.01	52	13	36.11	9.03
TSC_C08640	1488	0.49	TTC0920	1494	0.50	514	96%	0.01	279	72	54.28	14.01
TSC_C20300	237	0.47	TTC1124	237	0.48	78	99%	0.01	38	9	48.72	11.54
TSC_C01980	414	0.44	TTC1536	423	0.45	141	97%	0.01	95	30	67.38	21.28
TSC_C04970	828	0.55	TTC0985	843	0.56	287	96%	0.01	170	47	59.23	16.38
TSC_C20940	1932	0.50	TTC1038	1932	0.52	647	99%	0.01	279	58	43.12	8.96
TSC_C22040	522	0.51	TTC1210	528	0.52	178	97%	0.01	70	17	39.33	9.55
TSC_C18750	969	0.50	TTC1081	981	0.51	326	99%	0.01	104	12	31.90	3.68
TSC_C01960	867	0.51	TTC1539	879	0.53	294	98%	0.01	139	37	47.28	12.59
TSC_C08870	435	0.52	TTC0381	435	0.53	152	94%	0.01	115	48	75.66	31.58
TSC_C03080	1254	0.54	TTC1422	1251	0.55	426	97%	0.01	217	55	50.94	12.91
TSC_C16950	789	0.44	TTC0509	771	0.46	265	97%	0.01	81	19	30.57	7.17
TSC_C02150	1143	0.52	TTC1783	1140	0.53	390	97%	0.01	175	47	44.87	12.05
TSC_C13730	936	0.48	TTC1433	1161	0.50	410	82%	0.01	185	67	45.12	16.34
TSC_C17140	1377	0.47	TTC0489	1371	0.48	465	98%	0.01	210	55	45.16	11.83
TSC_C04490	360	0.42	TTC1974	357	0.43	121	97%	0.01	83	26	68.60	21.49
TSC_C06910	1425	0.49	TTC0151	1422	0.50	475	99%	0.01	178	46	37.47	9.68
TSC_C17920	780	0.52	TTC0729	816	0.53	272	97%	0.01	127	29	46.69	10.66
TSC_C16990	831	0.47	TTC0505	825	0.48	277	99%	0.01	144	43	51.99	15.52
TSC_C09110	1008	0.55	TTC0343	1008	0.56	344	97%	0.01	132	40	38.37	11.63
TSC_C12210	477	0.42	TTC0618	405	0.44	159	90%	0.01	58	16	36.48	10.06
TSC_C05970	549	0.44	TTC1912	549	0.45	183	99%	0.01	52	4	28.42	2.19
TSC_C06970	597	0.54	TTC0157	597	0.55	198	99%	0.01	68	17	34.34	8.59
TSC_C09950	399	0.44	TTC0270	387	0.45	136	95%	0.01	92	32	67.65	23.53
TSC_C00370	750	0.52	TTC1625	756	0.53	259	96%	0.01	134	32	51.74	12.36

TSC_C23180	1554	0.45	TTC1265	1551	0.46	527	98%	0.01	331	101	62.81	19.17
TSC_C06470	903	0.50	TTC1863	885	0.51	309	96%	0.01	204	72	66.02	23.30
TSC_C22010	1023	0.46	TTC1205	1038	0.47	345	99%	0.01	104	23	30.14	6.67
TSC_C20360	414	0.42	TTC1116	414	0.43	138	99%	0.01	66	15	47.83	10.87
TSC_C06050	360	0.49	TTC1920	360	0.50	120	98%	0.01	40	8	33.33	6.67
TSC_C07210	1035	0.50	TTC0448	1035	0.51	345	99%	0.01	141	27	40.87	7.83
TSC_C03910	1407	0.46	TTC1488	1407	0.47	469	100%	0.01	140	29	29.85	6.18
TSC_C21730	660	0.52	TTC1145	660	0.53	223	98%	0.01	122	38	54.71	17.04
TSC_C23240	381	0.46	TTC1272	381	0.47	126	99%	0.01	43	6	34.13	4.76
TSC_C13480	996	0.51	TTC0568	981	0.52	331	99%	0.01	165	21	49.85	6.34
TSC_C22350	318	0.43	TTC1319	318	0.44	105	99%	0.01	31	7	29.52	6.67
TSC_C19570	1056	0.52	TTC0117	1056	0.53	356	98%	0.01	175	36	49.16	10.11
TSC_C20340	426	0.47	TTC1120	414	0.48	141	98%	0.01	61	19	43.26	13.48
TSC_C18060	435	0.51	TTC0476	447	0.52	154	94%	0.01	107	34	69.48	22.08
TSC_C13960	759	0.50	TTC0259	759	0.51	252	100%	0.01	126	25	50.00	9.92
TSC_C01510	630	0.46	TTC1618	633	0.47	212	98%	0.01	155	46	73.11	21.70
TSC_C18210	1200	0.48	TTC0435	1158	0.49	400	98%	0.01	140	27	35.00	6.75
TSC_C02390	975	0.48	TTC1756	975	0.50	324	100%	0.01	94	10	29.01	3.09
TSC_C09850	1986	0.47	TTC0752	1986	0.48	667	99%	0.01	283	69	42.43	10.34
TSC_C22470	2076	0.47	TTC1331	2076	0.48	692	100%	0.01	172	21	24.86	3.03
TSC_C02790	1176	0.48	TTC1706	1176	0.49	391	100%	0.01	105	22	26.85	5.63
TSC_C07470	1545	0.47	TTC0882	1545	0.48	515	100%	0.01	185	31	35.92	6.02
TSC_C12180	456	0.51	TTC0605	513	0.52	172	92%	0.01	49	15	28.49	8.72
TSC_C08030	873	0.47	TTC0898	873	0.48	293	99%	0.01	118	23	40.27	7.85
TSC_C22140	834	0.47	TTC1297	834	0.48	290	95%	0.01	192	62	66.21	21.38
TSC_C19690	1020	0.47	TTC0101	1023	0.48	341	99%	0.01	125	25	36.66	7.33
TSC_C11840	1332	0.48	TTC0319	1308	0.49	445	98%	0.01	167	21	37.53	4.72
TSC_C12150	972	0.45	TTC0608	972	0.46	323	100%	0.01	94	15	29.10	4.64
TSC_C05480	759	0.55	TTC1073	714	0.56	253	96%	0.01	114	33	45.06	13.04

TSC_C22810	2514	0.50	TTC0028	2475	0.51	857	97%	0.01	459	146	53.56	17.04
TSC_C20350	687	0.46	TTC1119	687	0.47	228	100%	0.01	94	25	41.23	10.96
TSC_C15020	951	0.48	TTC0793	921	0.49	316	98%	0.01	114	23	36.08	7.28
TSC_C06340	1134	0.46	TTC0978	1134	0.47	377	100%	0.01	146	32	38.73	8.49
TSC_C02680	2733	0.51	TTC1698	2691	0.52	924	98%	0.01	373	109	40.37	11.80
TSC_C10470	957	0.47	TTC0596	960	0.48	328	97%	0.01	145	43	44.21	13.11
TSC_C01970	846	0.48	TTC1538	846	0.49	284	99%	0.01	110	27	38.73	9.51
TSC_C22450	318	0.42	TTC1329	318	0.43	105	99%	0.01	32	6	30.48	5.71
TSC_C22290	543	0.50	TTC1313	543	0.51	181	99%	0.01	73	17	40.33	9.39
TSC_C19070	1683	0.43	TTC1099	1683	0.44	563	99%	0.01	238	62	42.27	11.01
TSC_C03650	552	0.50	TTC1380	552	0.51	185	98%	0.01	77	11	41.62	5.95
TSC_C06890	387	0.44	TTC0149	387	0.45	129	98%	0.01	54	11	41.86	8.53
TSC_C21390	729	0.48	TTC1242	729	0.49	247	98%	0.01	109	30	44.13	12.15
TSC_C17190	660	0.49	TTC0484	657	0.50	221	98%	0.01	112	33	50.68	14.93
TSC_C02850	1107	0.49	TTC0744	1107	0.50	370	99%	0.01	199	65	53.78	17.57
TSC_C21510	324	0.45	TTC1253	324	0.46	108	98%	0.01	38	7	35.19	6.48
TSC_C19720	1752	0.43	TTC0098	1749	0.44	585	99%	0.01	226	54	38.63	9.23
TSC_C05930	1410	0.50	TTC1908	1410	0.51	473	99%	0.01	233	59	49.26	12.47
TSC_C08530	684	0.52	TTC0839	684	0.53	227	100%	0.01	102	20	44.93	8.81
TSC_C03120	375	0.45	TTC1418	342	0.46	124	95%	0.01	46	14	37.10	11.29
TSC_C23220	1011	0.46	TTC1270	1011	0.47	337	99%	0.01	132	20	39.17	5.93
TSC_C22410	831	0.49	TTC1325	831	0.50	276	100%	0.01	78	9	28.26	3.26
TSC_C15530	2055	0.54	TTC0354	2052	0.55	695	98%	0.01	217	67	31.22	9.64
TSC_C21610	651	0.46	TTC1137	618	0.47	219	96%	0.01	115	36	52.51	16.44
TSC_C03470	516	0.50	TTC1401	513	0.51	176	96%	0.01	95	36	53.98	20.45
TSC_C13940	537	0.48	TTC0692	543	0.49	181	98%	0.01	69	14	38.12	7.73
TSC_C22360	198	0.43	TTC1320	219	0.44	72	94%	0.01	32	12	44.44	16.67
TSC_C13400	486	0.43	TTC0560	483	0.44	161	99%	0.01	62	11	38.51	6.83
TSC_C02330	651	0.53	TTC1751	651	0.54	221	97%	0.01	90	23	40.72	10.41

TSC_C10000	543	0.53	TTC0265	567	0.54	188	97%	0.01	87	11	46.28	5.85
TSC_C15490	1083	0.44	TTC0414	1086	0.45	361	100%	0.01	235	55	65.10	15.24
TSC_C02400	1104	0.48	TTC1757	1104	0.49	373	98%	0.01	113	23	30.29	6.17
TSC_C24690	2652	0.48	TTC1825	2637	0.49	889	99%	0.01	314	77	35.32	8.66
TSC_C24170	771	0.45	TTC1528	771	0.46	259	98%	0.01	125	33	48.26	12.74
TSC_C24600	1329	0.52	TTC1950	1341	0.52	450	98%	0.01	236	65	52.44	14.44
TSC_C01590	474	0.48	TTC1647	474	0.49	157	99%	0.01	48	3	30.57	1.91
TSC_C18410	792	0.48	TTC1684	852	0.49	321	82%	0.01	278	104	86.60	32.40
TSC_C17210	849	0.51	TTC0234	849	0.51	284	99%	0.01	171	51	60.21	17.96
TSC_C23160	777	0.46	TTC1263	777	0.47	263	98%	0.01	126	45	47.91	17.11
TSC_C20490	4479	0.53	TTC1104	4479	0.54	1496	100%	0.01	544	101	36.36	6.75
TSC_C17840	561	0.44	TTC0927	561	0.44	186	99%	0.01	83	20	44.62	10.75
TSC_C10290	1035	0.37	TTC0582	1038	0.37	354	97%	0.01	129	37	36.44	10.45
TSC_C02060	1536	0.50	TTC1715	1533	0.51	514	99%	0.01	224	63	43.58	12.26
TSC_C22190	390	0.42	TTC1302	390	0.43	129	99%	0.01	41	5	31.78	3.88
TSC_C14970	738	0.52	TTC0800	765	0.53	259	96%	0.01	142	38	54.83	14.67
TSC_C04930	444	0.45	TTC0989	450	0.46	149	99%	0.01	68	12	45.64	8.05
TSC_C24330	576	0.48	TTC1662	579	0.49	192	99%	0.01	95	21	49.48	10.94
TSC_C12990	1077	0.48	TTC1790	1080	0.49	371	96%	0.01	241	84	64.96	22.64
TSC_C05240	2115	0.41	TTC0952	2124	0.42	729	97%	0.01	357	90	48.97	12.35
TSC_C20550	777	0.50	TTC1005	762	0.50	273	93%	0.01	161	56	58.97	20.51
TSC_C02240	267	0.46	TTC1742	267	0.46	88	99%	0.01	25	5	28.41	5.68
TSC_C06920	441	0.45	TTC0152	441	0.45	150	97%	0.01	84	30	56.00	20.00
TSC_C22830	684	0.52	TTC0026	684	0.52	229	99%	0.01	119	29	51.97	12.66
TSC_C04860	363	0.50	TTC0996	366	0.51	128	93%	0.01	101	34	78.91	26.56
TSC_C07290	591	0.53	TTC0456	534	0.54	204	90%	0.01	113	34	55.39	16.67
TSC_C07220	780	0.48	TTC0449	774	0.49	270	95%	0.01	155	36	57.41	13.33
TSC_C08310	357	0.46	TTC0185	357	0.47	118	99%	0.01	46	11	38.98	9.32
TSC_C08550	369	0.42	TTC0367	369	0.43	122	99%	0.01	44	6	36.07	4.92

TSC_C09070	426	0.48	TTC0339	426	0.49	142	99%	0.01	50	11	35.21	7.75
TSC_C03920	753	0.48	TTC1487	753	0.49	250	100%	0.01	79	16	31.60	6.40
TSC_C01390	876	0.43	TTC1628	876	0.43	291	100%	0.01	138	30	47.42	10.31
TSC_C18280	747	0.50	TTC0430	738	0.50	252	97%	0.01	159	45	63.10	17.86
TSC_C09620	462	0.48	TTC1692	435	0.49	169	86%	0.01	121	46	71.60	27.22
TSC_C08350	615	0.42	TTC0189	615	0.43	204	100%	0.01	62	12	30.39	5.88
TSC_C22200	381	0.48	TTC1303	381	0.48	126	99%	0.01	36	2	28.57	1.59
TSC_C20860	711	0.52	TTC1029	741	0.52	252	95%	0.01	135	43	53.57	17.06
TSC_C17350	654	0.47	TTC1138	684	0.48	252	86%	0.01	185	79	73.41	31.35
TSC_C07960	990	0.50	TTC0904	990	0.50	334	98%	0.01	163	52	48.80	15.57
TSC_C03810	516	0.44	TTC1362	519	0.45	179	95%	0.00	117	30	65.36	16.76
TSC_C12170	759	0.49	TTC0606	747	0.50	255	98%	0.00	126	39	49.41	15.29
TSC_C20990	669	0.49	TTC1044	669	0.49	222	100%	0.00	124	28	55.86	12.61
TSC_C05520	822	0.51	TTC1069	576	0.51	276	81%	0.00	131	42	47.46	15.22
TSC_C15570	1284	0.46	TTC0700	1281	0.47	427	100%	0.00	148	21	34.66	4.92
TSC_C24210	2664	0.47	TTC1532	2847	0.48	1062	84%	0.00	683	280	64.31	26.37
TSC_C13440	1266	0.47	TTC0408	1266	0.48	421	100%	0.00	163	29	38.72	6.89
TSC_C09570	1449	0.52	TTC1702	1389	0.53	576	78%	0.00	429	159	74.48	27.60
TSC_C04620	831	0.49	TTC1987	831	0.50	287	96%	0.00	157	48	54.70	16.72
TSC_C01840	792	0.47	TTC1551	792	0.48	264	99%	0.00	107	21	40.53	7.95
TSC_C03720	1215	0.51	TTC1373	1215	0.51	408	99%	0.00	167	37	40.93	9.07
TSC_C09640	2145	0.50	TTC0774	2142	0.50	720	99%	0.00	279	53	38.75	7.36
TSC_C19750	549	0.43	TTC1797	546	0.44	222	77%	0.00	149	69	67.12	31.08
TSC_C22380	720	0.49	TTC1322	720	0.49	239	100%	0.00	93	20	38.91	8.37
TSC_C00260	627	0.54	TTC1615	645	0.54	219	96%	0.00	144	51	65.75	23.29
TSC_C02000	699	0.52	TTC1534	699	0.52	236	98%	0.00	132	45	55.93	19.07
TSC_C17170	723	0.45	TTC0486	690	0.46	241	97%	0.00	77	20	31.95	8.30
TSC_C04230	3360	0.51	TTC1461	3360	0.52	1121	100%	0.00	345	73	30.78	6.51
TSC_C00360	207	0.39	TTC1626	207	0.40	68	99%	0.00	33	10	48.53	14.71

TSC_C01070	792	0.45	TTC1723	816	0.45	287	92%	0.00	203	78	70.73	27.18
TSC_C03500	612	0.49	TTC1398	612	0.49	203	100%	0.00	96	31	47.29	15.27
TSC_C00420	552	0.49	TTC1439	558	0.49	185	99%	0.00	64	17	34.59	9.19
TSC_C09260	1656	0.47	TTC0677	1656	0.47	551	100%	0.00	107	9	19.42	1.63
TSC_C06040	546	0.47	TTC1919	546	0.47	181	99%	0.00	47	2	25.97	1.10
TSC_C08420	1557	0.50	TTC0849	1563	0.50	521	99%	0.00	178	31	34.17	5.95
TSC_C08810	2706	0.49	TTC0374	2709	0.50	906	99%	0.00	290	67	32.01	7.40
TSC_C04910	312	0.50	TTC0990	312	0.50	103	99%	0.00	57	23	55.34	22.33
TSC_C03680	531	0.49	TTC1377	528	0.49	186	93%	0.00	86	17	46.24	9.14
TSC_C07630	606	0.47	TTC0866	606	0.47	201	100%	0.00	83	26	41.29	12.94
TSC_C01540	1089	0.44	TTC1621	1086	0.44	362	100%	0.00	140	25	38.67	6.91
TSC_C01940	810	0.47	TTC1541	810	0.47	270	99%	0.00	118	28	43.70	10.37
TSC_C21020	1290	0.48	TTC1047	1287	0.48	434	98%	0.00	202	51	46.54	11.75
TSC_C05300	627	0.47	TTC0946	627	0.47	208	100%	0.00	84	17	40.38	8.17
TSC_C08790	1437	0.45	TTC0906	1437	0.45	479	100%	0.00	103	6	21.50	1.25
TSC_C20150	705	0.52	TTC0092	705	0.52	236	99%	0.00	90	19	38.14	8.05
TSC_C24990	495	0.52	TTC1601	627	0.52	224	79%	0.00	110	39	49.11	17.41
TSC_C15180	360	0.43	TTC0820	360	0.43	122	97%	0.00	103	33	84.43	27.05
TSC_C02760	540	0.50	TTC1703	534	0.51	179	99%	0.00	75	19	41.90	10.61
TSC_C17040	786	0.55	TTC0500	786	0.56	262	99%	0.00	135	36	51.53	13.74
TSC_C00160	540	0.43	TTC1792	543	0.43	183	98%	0.00	115	34	62.84	18.58
TSC_C22130	525	0.49	TTC1296	513	0.49	176	97%	0.00	98	32	55.68	18.18
TSC_C22880	744	0.45	TTC0019	741	0.45	250	98%	0.00	96	24	38.40	9.60
TSC_C10230	1299	0.52	TTC0510	1299	0.52	435	99%	0.00	224	63	51.49	14.48
TSC_C02700	1374	0.53	TTC1700	1404	0.53	470	98%	0.00	193	49	41.06	10.43
TSC_C18200	210	0.46	TTC0891	210	0.46	70	97%	0.00	35	8	50.00	11.43
TSC_C06590	924	0.49	TTC1852	1137	0.49	380	89%	0.00	134	41	35.26	10.79
TSC_C13070	672	0.47	TTC0783	660	0.47	230	95%	0.00	117	31	50.87	13.48
TSC_C07680	1014	0.45	TTC0864	1014	0.45	339	99%	0.00	125	28	36.87	8.26

TSC_C14820	1110	0.43	TTC0326	1110	0.43	374	98%	0.00	166	42	44.39	11.23
TSC_C24200	2001	0.50	TTC1531	2007	0.50	674	99%	0.00	238	47	35.31	6.97
TSC_C12960	1473	0.51	TTC1915	1317	0.50	576	76%	0.00	379	159	65.80	27.60
TSC_C14720	672	0.51	TTC0398	660	0.51	227	97%	0.00	142	44	62.56	19.38
TSC_C07000	489	0.46	TTC0160	513	0.45	171	96%	0.00	89	21	52.05	12.28
TSC_C13090	249	0.47	TTC0786	240	0.47	84	94%	0.00	46	11	54.76	13.10
TSC_C21830	2208	0.51	TTC1154	2178	0.51	742	98%	0.00	319	76	42.99	10.24
TSC_C06780	1902	0.50	TTC0637	1827	0.50	646	96%	0.00	379	113	58.67	17.49
TSC_C05250	984	0.45	TTC0951	1050	0.45	349	96%	0.00	155	53	44.41	15.19
TSC_C22370	426	0.45	TTC1321	426	0.45	141	99%	0.00	40	6	28.37	4.26
TSC_C03690	711	0.50	TTC1376	711	0.50	237	99%	0.00	140	45	59.07	18.99
TSC_C18150	972	0.56	TTC0888	972	0.56	328	98%	0.00	180	51	54.88	15.55
TSC_C01930	1041	0.46	TTC1542	1035	0.46	347	99%	0.00	137	26	39.48	7.49
TSC_C13710	681	0.52	TTC0540	687	0.51	230	98%	0.00	104	21	45.22	9.13
TSC_C05950	288	0.48	TTC1910	288	0.48	95	99%	0.00	28	7	29.47	7.37
TSC_C03940	1317	0.48	TTC1485	1308	0.48	439	99%	0.00	133	26	30.30	5.92
TSC_C20410	834	0.41	TTC1111	828	0.41	279	99%	0.00	117	22	41.94	7.89
TSC_C17060	273	0.44	TTC0498	273	0.44	90	99%	0.00	17	2	18.89	2.22
TSC_C13600	753	0.50	TTC0581	753	0.50	250	100%	0.00	121	29	48.40	11.60
TSC_C05330	351	0.40	TTC0122	390	0.40	129	94%	0.00	42	5	32.56	3.88
TSC_C15030	777	0.45	TTC0806	786	0.45	266	97%	0.00	143	29	53.76	10.90
TSC_C06030	621	0.47	TTC1918	624	0.46	207	99%	0.00	83	23	40.10	11.11
TSC_C06940	588	0.46	TTC0154	588	0.46	195	99%	0.00	64	22	32.82	11.28
TSC_C02600	1290	0.53	TTC0646	1314	0.52	462	93%	0.00	357	127	77.27	27.49
TSC_C21590	918	0.52	TTC1135	891	0.51	306	98%	0.00	130	36	42.48	11.76
TSC_C08700	210	0.49	TTC0915	210	0.48	69	99%	0.00	36	12	52.17	17.39
TSC_C02380	210	0.42	TTC1755	204	0.41	70	96%	0.00	42	10	60.00	14.29
TSC_C03450	1353	0.44	TTC1403	1347	0.44	453	99%	0.00	155	47	34.22	10.38
TSC_C13530	753	0.45	TTC0575	753	0.44	250	100%	0.00	27	6	10.80	2.40

TSC_C08240	306	0.51	TTC0179	312	0.51	105	96%	-0.01	59	22	56.19	20.95
TSC_C03130	552	0.47	TTC0860	543	0.46	203	88%	-0.01	140	45	68.97	22.17
TSC_C08470	786	0.47	TTC0842	810	0.46	278	95%	-0.01	141	32	50.72	11.51
TSC_C01620	615	0.50	TTC1650	663	0.49	227	92%	-0.01	125	39	55.07	17.18
TSC_C08540	618	0.45	TTC0366	618	0.45	205	100%	-0.01	69	12	33.66	5.85
TSC_C05230	978	0.49	TTC0953	984	0.48	330	98%	-0.01	173	55	52.42	16.67
TSC_C20590	759	0.46	TTC1007	759	0.45	255	98%	-0.01	101	25	39.61	9.80
TSC_C17120	3192	0.45	TTC0491	2007	0.45	1065	77%	-0.01	221	49	20.75	4.60
TSC_C03760	654	0.46	TTC1368	636	0.46	224	95%	-0.01	103	23	45.98	10.27
TSC_C07170	483	0.45	TTC0442	468	0.45	180	85%	-0.01	155	69	86.11	38.33
TSC_C21370	267	0.40	TTC1240	267	0.39	88	99%	-0.01	33	10	37.50	11.36
TSC_C22440	621	0.48	TTC1328	621	0.48	206	100%	-0.01	83	15	40.29	7.28
TSC_C14740	738	0.52	TTC0394	738	0.52	248	98%	-0.01	112	19	45.16	7.66
TSC_C17220	1047	0.51	TTC0232	1083	0.50	367	96%	-0.01	169	53	46.05	14.44
TSC_C03290	222	0.42	TTC1811	258	0.41	85	91%	-0.01	20	3	23.53	3.53
TSC_C20380	420	0.47	TTC1114	459	0.46	156	92%	-0.01	71	16	45.51	10.26
TSC_C09080	732	0.54	TTC0340	732	0.53	247	98%	-0.01	151	53	61.13	21.46
TSC_C00170	1500	0.47	TTC1791	1500	0.46	502	99%	-0.01	254	55	50.60	10.96
TSC_C03870	285	0.48	TTC1357	285	0.47	95	98%	-0.01	37	6	38.95	6.32
TSC_C09650	270	0.43	TTC0773	270	0.42	89	99%	-0.01	41	10	46.07	11.24
TSC_C02650	765	0.54	TTC1697	765	0.53	254	100%	-0.01	128	33	50.39	12.99
TSC_C21750	579	0.49	TTC1146	582	0.48	197	97%	-0.01	112	35	56.85	17.77
TSC_C05790	960	0.49	TTC0041	927	0.48	337	92%	-0.01	153	40	45.40	11.87
TSC_C21640	372	0.46	TTC1140	261	0.45	123	82%	-0.01	35	7	28.46	5.69
TSC_C02740	1203	0.49	TTC1701	1203	0.48	402	99%	-0.01	131	25	32.59	6.22
TSC_C07500	804	0.47	TTC0878	804	0.46	267	100%	-0.01	129	28	48.31	10.49
TSC_C01790	690	0.45	TTC1739	690	0.44	229	100%	-0.01	77	11	33.62	4.80
TSC_C07240	813	0.49	TTC0451	816	0.48	281	96%	-0.01	153	56	54.45	19.93
TSC_C08590	792	0.50	TTC0371	798	0.48	266	99%	-0.01	135	36	50.75	13.53

TSC_C22890	276	0.44	TTC0018	222	0.43	92	87%	-0.01	26	8	28.26	8.70
TSC_C08600	1332	0.47	TTC0372	1377	0.45	494	90%	-0.01	424	145	85.83	29.35
TSC_C02320	1713	0.50	TTC1750	1332	0.48	577	86%	-0.01	207	49	35.88	8.49
TSC_C22610	669	0.44	TTC1765	669	0.43	223	99%	-0.01	96	26	43.05	11.66
TSC_C02340	279	0.49	TTC1752	273	0.48	92	98%	-0.01	42	12	45.65	13.04
TSC_C04510	1245	0.46	TTC1976	1245	0.44	416	99%	-0.01	141	24	33.89	5.77
TSC_C19600	339	0.44	TTC0120	339	0.43	118	94%	-0.01	85	23	72.03	19.49
TSC_C05850	738	0.54	TTC0047	738	0.52	247	99%	-0.01	95	18	38.46	7.29
TSC_C08520	603	0.51	TTC0841	600	0.49	206	96%	-0.01	110	23	53.40	11.17
TSC_C22260	183	0.40	TTC1310	183	0.39	60	98%	-0.01	33	12	55.00	20.00
TSC_C03730	2025	0.52	TTC1371	2064	0.50	707	96%	-0.01	478	147	67.61	20.79
TSC_C06650	1176	0.55	TTC0665	1176	0.53	392	99%	-0.02	57	16	14.54	4.08
TSC_C06500	462	0.48	TTC1860	462	0.47	155	98%	-0.02	67	13	43.23	8.39
TSC_C21130	447	0.47	TTC1058	447	0.45	149	99%	-0.02	67	22	44.97	14.77
TSC_C14280	381	0.46	TTC0261	375	0.44	131	94%	-0.02	87	32	66.41	24.43
TSC_C13700	249	0.41	TTC0541	249	0.39	82	99%	-0.02	43	15	52.44	18.29
TSC_C22480	471	0.47	TTC1332	471	0.45	156	99%	-0.02	54	6	34.62	3.85
TSC_C17730	696	0.49	TTC1089	699	0.47	239	96%	-0.02	156	62	65.27	25.94
TSC_C16910	993	0.56	TTC0656	882	0.54	343	90%	-0.02	222	84	64.72	24.49
TSC_C21440	228	0.36	TTC1246	237	0.34	80	94%	-0.02	20	5	25.00	6.25
TSC_C22330	333	0.44	TTC1317	333	0.42	110	99%	-0.02	44	14	40.00	12.73
TSC_C05840	243	0.39	TTC0046	243	0.37	80	99%	-0.02	21	4	26.25	5.00
TSC_C02920	1122	0.43	TTC0529	1119	0.41	376	99%	-0.02	209	60	55.59	15.96
TSC_C24700	783	0.47	TTC1566	819	0.45	284	93%	-0.02	176	54	61.97	19.01
TSC_C14780	855	0.48	TTC0390	855	0.46	284	100%	-0.02	140	36	49.30	12.68
TSC_C02530	615	0.52	TTC0640	615	0.49	211	96%	-0.02	150	55	71.09	26.07
TSC_C03430	636	0.48	TTC1406	639	0.45	217	97%	-0.02	105	31	48.39	14.29
TSC_C03570	378	0.45	TTC1386	429	0.43	143	92%	-0.02	51	10	35.66	6.99
TSC_C03880	201	0.41	TTC1356	201	0.38	67	97%	-0.02	36	12	53.73	17.91

TSC_C02570	1203	0.58	TTC0644	1218	0.56	421	95%	-0.02	244	71	57.96	16.86
TSC_C18130	735	0.48	TTC0469	735	0.46	244	100%	-0.02	99	23	40.57	9.43
TSC_C15880	1053	0.44	TTC1010	1413	0.42	504	77%	-0.02	333	133	66.07	26.39
TSC_C06250	303	0.40	TTC0140	303	0.38	100	99%	-0.02	52	11	52.00	11.00
TSC_C22730	387	0.49	TTC1221	384	0.47	130	97%	-0.03	77	14	59.23	10.77
TSC_C13850	462	0.43	TTC0669	441	0.41	157	94%	-0.03	73	25	46.50	15.92
TSC_C20790	633	0.49	TTC1023	570	0.46	216	91%	-0.03	120	33	55.56	15.28
TSC_C24510	219	0.43	TTC1365	213	0.40	78	89%	-0.03	80	28	102.56	35.90
TSC_C06720	2049	0.54	TTC0690	2505	0.51	950	75%	-0.03	627	252	66.00	26.53
TSC_C07190	690	0.53	TTC1827	654	0.50	241	92%	-0.03	132	45	54.77	18.67
TSC_C14110	330	0.48	TTC0246	330	0.45	110	98%	-0.03	48	13	43.64	11.82
TSC_C13570	1464	0.63	TTC0579	1464	0.60	491	99%	-0.03	84	29	17.11	5.91
TSC_C03840	213	0.46	TTC1359	213	0.43	71	97%	-0.03	25	7	35.21	9.86
TSC_C02500	1056	0.53	TTC1353	891	0.50	364	87%	-0.03	164	58	45.05	15.93
TSC_C02180	1128	0.50	TTC1779	1128	0.47	375	100%	-0.03	122	29	32.53	7.73
TSC_C10770	825	0.49	TTC0973	849	0.46	307	89%	-0.03	256	93	83.39	30.29
TSC_C05980	1119	0.44	TTC1913	1098	0.41	375	98%	-0.03	146	36	38.93	9.60
TSC_C22520	429	0.48	TTC1774	432	0.44	149	95%	-0.04	91	32	61.07	21.48
TSC_C09090	774	0.54	TTC0341	774	0.50	261	98%	-0.04	171	54	65.52	20.69
TSC_C00430	321	0.48	TTC1440	321	0.44	106	99%	-0.04	60	13	56.60	12.26
TSC_C03700	204	0.46	TTC1375	207	0.42	71	93%	-0.05	34	11	47.89	15.49
TSC_C06750	747	0.45	TTC0653	666	0.40	280	80%	-0.05	251	98	89.64	35.00
TSC_C01080	870	0.46	TTC1725	903	0.40	358	78%	-0.06	275	118	76.82	32.96
TSC_C20960	687	0.45	TTC1040	687	0.35	230	99%	-0.10	95	24	41.30	10.43
TSC_C14730	1167	0.51	TTC0395	1167	0.37	400	97%	-0.14	273	87	68.25	21.75
											0.30	0.38

NOTES: IDs of proteins, for which 3D structures were constructed, are highlighted by yellow cell background.

*Top 100 MFE values are highlighted by grey cell background.

[†]Frequencies of nucleotide substitutions per 300 nucleotides and amino acid substitutions per 100 amino acids.

^{*}Correlation coefficients between Δ MFE values and normalized frequencies of nucleotide and amino acid substitutions.

Table 6.4.2: MFE values and residue substitution statistics calculated for orthologous genes of *T. scotoductus* SA-01 and *T. thermophilus* HB8. Table entries are ordered by Δ MFE values (the difference between MFE values calculated for *T. thermophilus* HB8 and *T. scotoductus* SA-01 orthologous genes).

<i>T. scotoductus</i> SA-01			<i>T. thermophilus</i> HB8			Protein alignment		Δ MFE (kcal/Mol)	Total substitutions		Normalized substitutions [†]	
Gene ID	Length	MFE* (kcal/Mol)	Gene ID	Length	MFE* (kcal/Mol)	Length	Coverage		DNA	AMC	DNA	AMC
TSC_C18260	945	0.34	TTHA0786	936	0.53	317	98%	1.87	166	40	52.37	12.62
TSC_C18350	537	0.32	TTHA0778	522	0.50	204	83%	1.73	164	63	80.39	30.88
TSC_C15270	528	0.46	TTHA1192	540	0.63	183	96%	1.71	127	36	69.40	19.67
TSC_C22550	342	0.45	TTHA0215	342	0.61	116	96%	1.62	74	25	63.79	21.55
TSC_C13260	1038	0.39	TTHA0887	1050	0.54	359	96%	1.56	303	90	84.40	25.07
TSC_C11420	711	0.40	TTHA0869	711	0.55	291	76%	1.53	242	94	83.16	32.30
TSC_C08020	1044	0.43	TTHA0157	1050	0.58	352	99%	1.52	164	41	46.59	11.65
TSC_C18820	555	0.46	TTHA0375	549	0.61	195	93%	1.51	104	30	53.33	15.38
TSC_C01750	165	0.26	TTHA0249	165	0.41	54	98%	1.51	23	7	42.59	12.96
TSC_C05390	387	0.37	TTHA0501	393	0.52	135	95%	1.50	125	33	92.59	24.44
TSC_C00780	306	0.39	TTHA0181	303	0.53	110	90%	1.39	102	34	92.73	30.91
TSC_C08560	1140	0.49	TTHA0720	1065	0.63	404	90%	1.39	308	101	76.24	25.00
TSC_C01170	315	0.44	TTHA0254	285	0.58	110	88%	1.39	79	27	71.82	24.55
TSC_C13360	1110	0.48	TTHA0871	1110	0.61	374	98%	1.32	236	77	63.10	20.59
TSC_C13580	960	0.43	TTHA0153	960	0.56	319	100%	1.30	151	38	47.34	11.91
TSC_C13250	582	0.36	TTHA0888	579	0.49	204	94%	1.30	153	42	75.00	20.59
TSC_C00510	819	0.45	TTHA0992	924	0.58	336	84%	1.29	232	79	69.05	23.51

TSC_C02970	1125	0.41	TTHA0643	1095	0.54	457	76%	1.27	376	147	82.28	32.17
TSC_C07380	1749	0.48	TTHA0817	1758	0.61	615	94%	1.25	446	144	72.52	23.41
TSC_C21040	363	0.39	TTHA1414	435	0.51	146	89%	1.25	61	13	41.78	8.90
TSC_C19440	213	0.41	TTHA0592	189	0.54	73	88%	1.24	53	19	72.60	26.03
TSC_C07880	615	0.49	TTHA1071	618	0.61	214	95%	1.24	158	46	73.83	21.50
TSC_C00480	681	0.51	TTHA0071	675	0.63	234	96%	1.23	183	63	78.21	26.92
TSC_C01450	807	0.50	TTHA0350	807	0.62	278	96%	1.20	210	64	75.54	23.02
TSC_C19830	1050	0.49	TTHA1439	915	0.61	363	89%	1.19	249	94	68.60	25.90
TSC_C21460	630	0.46	TTHA1612	612	0.58	216	95%	1.18	150	43	69.44	19.91
TSC_C10340	444	0.46	TTHA0953	435	0.57	152	95%	1.16	117	41	76.97	26.97
TSC_C00730	462	0.46	TTHA1954	468	0.58	158	97%	1.15	101	34	63.92	21.52
TSC_C09160	279	0.43	TTHA0700	276	0.54	92	98%	1.14	58	21	63.04	22.83
TSC_C15130	741	0.49	TTHA1179	789	0.60	270	93%	1.13	153	45	56.67	16.67
TSC_C20070	366	0.39	TTHA0113	366	0.50	122	98%	1.12	63	22	51.64	18.03
TSC_C20740	372	0.36	TTHA0487	393	0.47	141	88%	1.12	122	44	86.52	31.21
TSC_C14930	564	0.42	TTHA1168	564	0.53	189	98%	1.11	111	37	58.73	19.58
TSC_C15120	762	0.51	TTHA1178	753	0.62	271	92%	1.11	172	48	63.47	17.71
TSC_C24830	822	0.49	TTHA0319	822	0.60	277	98%	1.10	166	41	59.93	14.80
TSC_C05570	897	0.51	TTHA1428	888	0.62	317	93%	1.09	217	73	68.45	23.03
TSC_C24760	285	0.42	TTHA1925	285	0.53	94	99%	1.08	39	10	41.49	10.64
TSC_C20690	945	0.40	TTHA1382	879	0.50	340	88%	1.08	247	78	72.65	22.94
TSC_C17230	924	0.49	TTHA0602	1014	0.59	358	89%	1.08	202	70	56.42	19.55
TSC_C00020	1116	0.46	FTTHA0001	1128	0.56	398	93%	1.08	253	84	63.57	21.11
TSC_C08840	1191	0.51	TTHA0730	1179	0.61	411	95%	1.08	275	95	66.91	23.11
TSC_C00380	1449	0.49	TTHA0338	1440	0.60	530	90%	1.08	382	137	72.08	25.85
TSC_C24130	522	0.49	TTHA1885	570	0.59	208	85%	1.07	157	63	75.48	30.29
TSC_C14180	591	0.46	TTHA0583	591	0.57	199	98%	1.07	103	30	51.76	15.08
TSC_C23850	390	0.47	TTHA1712	390	0.57	133	96%	1.06	81	29	60.90	21.80
TSC_C01040	762	0.48	TTHA0265	786	0.59	272	94%	1.04	170	67	62.50	24.63

TSC_C18010	552	0.47	TTHA1102	546	0.57	189	96%	1.04	112	34	59.26	17.99
TSC_C20670	624	0.43	TTHA1380	624	0.54	210	98%	1.03	143	44	68.10	20.95
TSC_C21150	612	0.49	TTHA1425	612	0.59	206	98%	1.03	101	27	49.03	13.11
TSC_C15540	360	0.46	TTHA0705	372	0.56	125	96%	1.02	62	11	49.60	8.80
TSC_C22790	1011	0.47	TTHA0398	1011	0.57	342	98%	1.01	216	73	63.16	21.35
TSC_C09920	498	0.46	TTHA0652	495	0.56	170	96%	0.99	122	41	71.76	24.12
TSC_C14420	435	0.41	TTHA1075	435	0.51	144	99%	0.98	61	13	42.36	9.03
TSC_C21960	1098	0.52	TTHA1565	957	0.62	376	90%	0.97	230	63	61.17	16.76
TSC_C06860	945	0.53	TTHA0521	882	0.63	329	91%	0.97	214	58	65.05	17.63
TSC_C07930	2931	0.52	TTHA1288	2901	0.62	1083	88%	0.97	723	245	66.76	22.62
TSC_C00450	273	0.44	TTHA0068	285	0.54	98	92%	0.97	78	31	79.59	31.63
TSC_C08690	372	0.45	TTHA1282	354	0.55	135	87%	0.97	103	31	76.30	22.96
TSC_C11740	753	0.48	TTHA0667	723	0.57	255	96%	0.96	162	52	63.53	20.39
TSC_C07920	1134	0.50	TTHA1289	1119	0.59	385	97%	0.96	252	83	65.45	21.56
TSC_C01630	270	0.41	TTHA0332	267	0.51	95	92%	0.96	45	17	47.37	17.89
TSC_C01050	1023	0.51	TTHA0264	1032	0.60	350	97%	0.96	233	70	66.57	20.00
TSC_C23760	927	0.51	TTHA1704	783	0.61	316	88%	0.96	191	62	60.44	19.62
TSC_C14480	858	0.47	TTHA1081	858	0.57	294	97%	0.96	158	51	53.74	17.35
TSC_C20680	624	0.40	TTHA1381	582	0.49	211	94%	0.95	145	43	68.72	20.38
TSC_C04780	3570	0.41	TTHA1020	3333	0.51	1368	81%	0.95	1036	379	75.73	27.70
TSC_C02930	909	0.37	TTHA0885	873	0.47	350	81%	0.95	269	87	76.86	24.86
TSC_C08850	969	0.47	TTHA0731	972	0.57	336	96%	0.94	247	84	73.51	25.00
TSC_C04250	579	0.48	TTHA1811	603	0.57	202	96%	0.94	120	29	59.41	14.36
TSC_C18460	336	0.47	TTHA1307	333	0.56	120	91%	0.94	78	21	65.00	17.50
TSC_C16780	750	0.49	TTHA0798	765	0.59	265	94%	0.94	155	55	58.49	20.75
TSC_C04890	231	0.40	TTHA1358	231	0.49	81	92%	0.94	56	16	69.14	19.75
TSC_C13680	3033	0.51	TTHA0899	3027	0.60	1082	93%	0.94	801	252	74.03	23.29
TSC_C24250	1236	0.51	TTHA1919	885	0.61	421	80%	0.94	192	51	45.61	12.11
TSC_C12950	1044	0.48	TTHA0595	1044	0.58	353	98%	0.93	217	55	61.47	15.58

TSC_C07100	453	0.47	TTHA1258	402	0.56	165	83%	0.93	94	33	56.97	20.00
TSC_C08890	351	0.47	TTHA0735	375	0.56	128	93%	0.93	63	14	49.22	10.94
TSC_C23320	774	0.49	TTHA1643	774	0.58	260	98%	0.93	144	38	55.38	14.62
TSC_C19340	1512	0.47	TTHA1552	1512	0.56	508	99%	0.92	233	47	45.87	9.25
TSC_C11850	777	0.49	TTHA0680	777	0.58	259	99%	0.92	164	46	63.32	17.76
TSC_C02880	414	0.51	TTHA1105	402	0.61	145	92%	0.92	113	37	77.93	25.52
TSC_C06060	819	0.44	TTHA0083	816	0.53	278	97%	0.92	161	51	57.91	18.35
TSC_C19730	1032	0.45	TTHA0466	1032	0.55	345	99%	0.91	193	37	55.94	10.72
TSC_C01650	618	0.49	TTHA0330	597	0.58	215	93%	0.91	136	45	63.26	20.93
TSC_C14750	315	0.47	TTHA0749	333	0.56	113	94%	0.91	61	13	53.98	11.50
TSC_C04260	225	0.46	TTHA1810	240	0.55	82	92%	0.91	40	15	48.78	18.29
TSC_C05220	423	0.49	TTHA1318	399	0.58	149	90%	0.91	107	39	71.81	26.17
TSC_C21540	615	0.48	TTHA0076	612	0.57	210	96%	0.91	150	41	71.43	19.52
TSC_C09120	2583	0.54	TTHA0696	2109	0.63	889	86%	0.91	487	115	54.78	12.94
TSC_C14320	831	0.45	TTHA0576	990	0.54	338	88%	0.90	135	39	39.94	11.54
TSC_C09700	588	0.38	TTHA1132	591	0.47	196	99%	0.90	91	16	46.43	8.16
TSC_C02900	1629	0.50	TTHA0830	1620	0.59	577	93%	0.90	438	162	75.91	28.08
TSC_C07110	2238	0.48	TTHA1257	2187	0.57	755	97%	0.90	430	135	56.95	17.88
TSC_C23810	1437	0.51	TTHA1709	1425	0.60	506	94%	0.90	330	101	65.22	19.96
TSC_C24960	498	0.48	TTHA1964	495	0.57	169	97%	0.90	105	38	62.13	22.49
TSC_C18610	366	0.44	TTHA1544	396	0.53	137	91%	0.89	109	30	79.56	21.90
TSC_C21100	327	0.38	TTHA1420	327	0.47	108	99%	0.89	52	16	48.15	14.81
TSC_C24880	2127	0.51	TTHA0063	2112	0.60	725	97%	0.89	497	131	68.55	18.07
TSC_C07810	453	0.43	TTHA0564	453	0.51	151	99%	0.89	98	39	64.90	25.83
TSC_C06310	435	0.39	TTHA1495	411	0.48	146	95%	0.89	68	20	46.58	13.70
TSC_C19820	330	0.43	TTHA0510	330	0.52	111	97%	0.88	67	17	60.36	15.32
TSC_C06330	1146	0.48	TTHA1493	1158	0.57	386	99%	0.88	249	66	64.51	17.10
TSC_C01580	726	0.50	TTHA0362	729	0.59	245	98%	0.88	138	32	56.33	13.06
TSC_C14800	1194	0.53	TTHA0689	1074	0.62	413	90%	0.88	253	84	61.26	20.34

TSC_C06320	933	0.48	TTHA1494	906	0.57	312	98%	0.88	220	71	70.51	22.76
TSC_C00880	2298	0.50	TTHA0276	2121	0.59	775	95%	0.87	388	115	50.06	14.84
TSC_C18340	1044	0.48	TTHA0779	1041	0.57	355	97%	0.87	208	52	58.59	14.65
TSC_C24870	1101	0.54	TTHA0062	1083	0.62	384	94%	0.87	212	61	55.21	15.89
TSC_C15340	522	0.40	TTHA1198	516	0.49	179	95%	0.87	156	45	87.15	25.14
TSC_C23280	570	0.46	TTHA1639	561	0.54	204	91%	0.87	156	49	76.47	24.02
TSC_C00870	774	0.50	TTHA0275	783	0.59	264	97%	0.87	168	51	63.64	19.32
TSC_C03890	303	0.42	TTHA1841	303	0.51	100	99%	0.87	40	9	40.00	9.00
TSC_C21350	330	0.51	TTHA1602	342	0.60	116	95%	0.86	67	32	57.76	27.59
TSC_C01190	858	0.50	TTHA0186	858	0.59	291	98%	0.86	152	47	52.23	16.15
TSC_C03180	519	0.41	TTHA1772	561	0.49	189	94%	0.85	129	45	68.25	23.81
TSC_C07840	8085	0.56	TTHA0568	8019	0.65	2874	93%	0.85	2144	738	74.60	25.68
TSC_C19770	534	0.47	TTHA0463	531	0.55	177	99%	0.85	103	31	58.19	17.51
TSC_C24550	648	0.48	TTHA0050	684	0.56	234	94%	0.85	134	36	57.26	15.38
TSC_C19330	468	0.47	TTHA1551	468	0.56	156	99%	0.85	79	20	50.64	12.82
TSC_C09240	489	0.45	TTHA1037	474	0.54	165	96%	0.85	67	19	40.61	11.52
TSC_C19560	999	0.47	TTHA0489	999	0.55	335	99%	0.85	193	47	57.61	14.03
TSC_C08440	534	0.41	TTHA1207	525	0.50	181	96%	0.84	96	28	53.04	15.47
TSC_C23030	2550	0.50	TTHA0073	2493	0.58	885	94%	0.84	436	114	49.27	12.88
TSC_C14330	1110	0.51	TTHA0575	1053	0.59	377	95%	0.84	223	63	59.15	16.71
TSC_C20660	1134	0.45	TTHA1379	1134	0.54	380	99%	0.84	253	81	66.58	21.32
TSC_C02410	414	0.48	TTHA0228	414	0.56	142	96%	0.84	76	19	53.52	13.38
TSC_C17390	954	0.48	TTHA0791	1083	0.57	396	83%	0.84	284	109	71.72	27.53
TSC_C21400	600	0.48	TTHA1607	597	0.56	209	94%	0.84	123	34	58.85	16.27
TSC_C05690	258	0.37	TTHA0444	249	0.45	87	95%	0.84	58	15	66.67	17.24
TSC_C00200	978	0.51	TTHA1787	921	0.59	368	83%	0.84	226	71	61.41	19.29
TSC_C23920	237	0.45	TTHA0298	237	0.54	78	99%	0.83	53	15	67.95	19.23
TSC_C00770	462	0.38	TTHA0182	447	0.47	170	86%	0.83	139	46	81.76	27.06
TSC_C11790	1731	0.45	TTHA0672	1758	0.54	587	99%	0.83	285	60	48.55	10.22

TSC_C14000	819	0.50	TTHA0621	741	0.58	281	91%	0.83	162	47	57.65	16.73
TSC_C19390	486	0.42	TTHA1556	501	0.51	178	91%	0.83	101	35	56.74	19.66
TSC_C15110	570	0.49	TTHA1177	846	0.57	281	80%	0.82	105	25	37.37	8.90
TSC_C08570	600	0.51	TTHA0721	567	0.59	207	93%	0.82	127	43	61.35	20.77
TSC_C22060	1260	0.51	TTHA1577	1260	0.59	429	97%	0.81	253	76	58.97	17.72
TSC_C12700	1992	0.50	TTHA0986	1929	0.58	684	95%	0.81	438	139	64.04	20.32
TSC_C18180	1332	0.50	TTHA1253	1317	0.58	461	95%	0.81	257	69	55.75	14.97
TSC_C14500	1059	0.48	TTHA1083	1068	0.56	391	89%	0.81	257	82	65.73	20.97
TSC_C13350	783	0.45	TTHA0873	783	0.53	262	99%	0.81	135	33	51.53	12.60
TSC_C04520	687	0.49	TTHA0021	687	0.57	228	100%	0.81	134	35	58.77	15.35
TSC_C20980	567	0.48	TTHA1408	570	0.56	195	96%	0.81	115	33	58.97	16.92
TSC_C09710	492	0.43	TTHA1131	492	0.51	164	99%	0.80	87	24	53.05	14.63
TSC_C05710	147	0.34	TTHA0446	150	0.42	50	95%	0.80	23	2	46.00	4.00
TSC_C24480	696	0.52	TTHA0041	582	0.60	232	90%	0.80	121	41	52.16	17.67
TSC_C20190	1326	0.53	TTHA0457	1284	0.61	449	96%	0.79	243	50	54.12	11.14
TSC_C10970	420	0.48	TTHA0895	414	0.56	162	82%	0.79	136	51	83.95	31.48
TSC_C18450	912	0.52	TTHA1306	870	0.60	313	94%	0.79	193	52	61.66	16.61
TSC_C06350	810	0.46	TTHA1344	816	0.54	274	98%	0.79	132	26	48.18	9.49
TSC_C04720	804	0.46	TTHA1024	783	0.54	290	90%	0.79	219	91	75.52	31.38
TSC_C14840	909	0.53	TTHA0664	909	0.61	308	98%	0.79	207	61	67.21	19.81
TSC_C20830	759	0.48	TTHA1392	765	0.56	262	96%	0.79	151	34	57.63	12.98
TSC_C19280	969	0.47	TTHA1587	960	0.55	322	99%	0.79	204	55	63.35	17.08
TSC_C04900	423	0.42	TTHA1357	426	0.50	141	99%	0.78	98	23	69.50	16.31
TSC_C04840	1251	0.50	TTHA1363	1254	0.58	427	97%	0.78	265	83	62.06	19.44
TSC_C21030	852	0.47	TTHA1413	855	0.55	287	98%	0.78	187	62	65.16	21.60
TSC_C09800	330	0.47	TTHA1122	339	0.55	117	93%	0.78	84	39	71.79	33.33
TSC_C07690	702	0.50	TTHA1226	657	0.58	241	93%	0.78	174	51	72.20	21.16
TSC_C05660	1386	0.56	TTHA0440	1344	0.63	491	92%	0.78	322	118	65.58	24.03
TSC_C15420	894	0.50	TTHA0710	894	0.58	301	98%	0.78	187	59	62.13	19.60

TSC_C23710	279	0.37	TTHA1700	273	0.44	92	98%	0.78	36	9	39.13	9.78
TSC_C06120	1896	0.51	TTHA0421	1887	0.59	649	97%	0.78	402	132	61.94	20.34
TSC_C18620	633	0.53	TTHA1545	540	0.61	217	88%	0.77	106	29	48.85	13.36
TSC_C04210	639	0.48	TTHA1814	642	0.56	219	96%	0.77	143	44	65.30	20.09
TSC_C00760	561	0.47	TTHA0183	564	0.55	191	97%	0.77	105	29	54.97	15.18
TSC_C14870	687	0.51	TTHA0657	687	0.59	232	98%	0.77	124	29	53.45	12.50
TSC_C21450	543	0.47	TTHA1611	555	0.54	199	90%	0.77	122	40	61.31	20.10
TSC_C01270	1062	0.43	TTHA1800	1095	0.51	378	94%	0.77	226	67	59.79	17.72
TSC_C02870	1440	0.53	TTHA1106	1434	0.61	511	93%	0.77	339	117	66.34	22.90
TSC_C14470	708	0.51	TTHA1080	699	0.58	257	90%	0.77	187	69	72.76	26.85
TSC_C21970	1455	0.51	TTHA1566	1455	0.58	504	96%	0.76	349	101	69.25	20.04
TSC_C02100	291	0.46	TTHA1793	327	0.54	110	91%	0.76	48	15	43.64	13.64
TSC_C08860	1269	0.44	TTHA0732	1266	0.52	444	94%	0.76	319	101	71.85	22.75
TSC_C08670	828	0.48	TTHA1284	615	0.55	280	83%	0.76	155	42	55.36	15.00
TSC_C11730	1050	0.45	TTHA0666	1050	0.53	354	98%	0.76	209	56	59.04	15.82
TSC_C20220	873	0.50	TTHA1346	873	0.57	295	98%	0.76	170	54	57.63	18.31
TSC_C15710	2487	0.53	TTHA0831	2457	0.61	848	97%	0.76	563	174	66.39	20.52
TSC_C19420	396	0.42	TTHA1528	396	0.49	132	98%	0.76	83	27	62.88	20.45
TSC_C21870	1605	0.53	TTHA1525	1587	0.60	563	94%	0.75	380	128	67.50	22.74
TSC_C04530	675	0.47	TTHA0020	672	0.54	227	98%	0.75	158	41	69.60	18.06
TSC_C21170	1086	0.47	TTHA1467	1038	0.54	372	94%	0.75	217	60	58.33	16.13
TSC_C00440	981	0.52	TTHA0067	981	0.60	339	96%	0.75	188	57	55.46	16.81
TSC_C07090	2652	0.53	TTHA1259	2616	0.61	955	91%	0.75	619	228	64.82	23.87
TSC_C19580	315	0.43	TTHA0492	309	0.50	107	95%	0.75	54	18	50.47	16.82
TSC_C22940	369	0.44	TTHA0372	339	0.52	124	93%	0.75	79	24	63.71	19.35
TSC_C23860	750	0.54	TTHA0292	750	0.61	267	92%	0.74	152	44	56.93	16.48
TSC_C23930	909	0.55	TTHA0299	909	0.62	308	98%	0.74	159	48	51.62	15.58
TSC_C24950	819	0.49	TTHA1963	795	0.56	278	96%	0.74	143	30	51.44	10.79
TSC_C04830	462	0.45	TTHA0434	486	0.52	166	94%	0.74	110	37	66.27	22.29

TSC_C12920	1671	0.52	TTHA0597	1662	0.60	569	97%	0.74	334	92	58.70	16.17
TSC_C24850	705	0.51	TTHA0317	705	0.58	249	93%	0.74	165	49	66.27	19.68
TSC_C25040	753	0.50	TTHA1970	750	0.57	250	99%	0.74	128	36	51.20	14.40
TSC_C07860	1062	0.50	TTHA0570	1059	0.58	356	99%	0.74	201	52	56.46	14.61
TSC_C17250	1356	0.56	TTHA0599	1374	0.63	468	97%	0.73	270	80	57.69	17.09
TSC_C21210	2619	0.54	TTHA0998	2574	0.61	919	94%	0.73	561	184	61.04	20.02
TSC_C13840	621	0.46	TTHA1030	579	0.54	212	93%	0.73	135	44	63.68	20.75
TSC_C20520	570	0.46	TTHA1367	564	0.53	197	95%	0.73	136	45	69.04	22.84
TSC_C20480	558	0.47	TTHA1469	612	0.54	206	93%	0.73	110	30	53.40	14.56
TSC_C05900	984	0.50	TTHA1496	876	0.57	333	92%	0.73	170	40	51.05	12.01
TSC_C08080	1278	0.53	TTHA0541	1272	0.60	432	98%	0.73	221	59	51.16	13.66
TSC_C21570	387	0.47	TTHA1497	447	0.55	149	91%	0.73	91	26	61.07	17.45
TSC_C23570	1224	0.47	TTHA0314	1221	0.54	438	92%	0.72	300	115	68.49	26.26
TSC_C19150	1038	0.51	TTHA0405	1032	0.58	351	98%	0.72	216	57	61.54	16.24
TSC_C11820	786	0.49	TTHA0676	807	0.57	271	97%	0.72	147	33	54.24	12.18
TSC_C05120	441	0.49	TTHA1327	444	0.56	148	98%	0.72	81	22	54.73	14.86
TSC_C07990	2328	0.48	TTHA1266	2313	0.55	781	99%	0.72	375	114	48.02	14.60
TSC_C12160	522	0.47	TTHA0971	522	0.54	179	96%	0.72	81	23	45.25	12.85
TSC_C19060	627	0.50	TTHA1462	627	0.57	211	98%	0.72	118	35	55.92	16.59
TSC_C10960	1170	0.46	TTHA1472	1161	0.53	476	77%	0.72	348	142	73.11	29.83
TSC_C20770	378	0.42	TTHA1388	378	0.49	127	98%	0.71	81	27	63.78	21.26
TSC_C10580	1140	0.51	TTHA0661	1200	0.58	402	96%	0.71	256	71	63.68	17.66
TSC_C14890	597	0.50	TTHA0655	603	0.57	202	98%	0.71	141	42	69.80	20.79
TSC_C23460	453	0.45	TTHA1853	453	0.52	151	99%	0.71	87	31	57.62	20.53
TSC_C08710	1146	0.46	TTHA1280	1149	0.53	390	98%	0.71	231	67	59.23	17.18
TSC_C20040	1215	0.51	TTHA0110	1212	0.58	408	99%	0.71	211	49	51.72	12.01
TSC_C18930	300	0.47	TTHA1451	300	0.54	100	98%	0.71	56	19	56.00	19.00
TSC_C00060	1470	0.51	TTHA1786	1464	0.58	504	97%	0.71	342	111	67.86	22.02
TSC_C06090	345	0.42	TTHA0080	339	0.50	120	93%	0.71	51	19	42.50	15.83

TSC_C08940	984	0.51	TTHA0740	978	0.58	337	96%	0.71	184	53	54.60	15.73
TSC_C20320	390	0.50	TTHA1486	390	0.57	132	97%	0.71	70	17	53.03	12.88
TSC_C23870	384	0.49	TTHA0293	369	0.56	127	97%	0.70	66	16	51.97	12.60
TSC_C04850	684	0.49	TTHA1362	684	0.56	228	99%	0.70	112	24	49.12	10.53
TSC_C22950	1353	0.55	TTHA0371	1320	0.62	469	94%	0.70	331	107	70.58	22.81
TSC_C03590	987	0.53	TTHA1746	981	0.60	333	98%	0.70	191	49	57.36	14.71
TSC_C23530	723	0.51	TTHA1861	723	0.58	242	99%	0.70	129	43	53.31	17.77
TSC_C07490	309	0.41	TTHA1244	306	0.48	103	98%	0.70	79	27	76.70	26.21
TSC_C15300	573	0.49	TTHA1195	576	0.56	203	93%	0.70	165	56	81.28	27.59
TSC_C14520	1341	0.51	TTHA1085	1341	0.58	464	96%	0.70	324	87	69.83	18.75
TSC_C09330	996	0.43	TTHA1046	996	0.50	333	99%	0.70	174	36	52.25	10.81
TSC_C14710	909	0.46	TTHA0758	912	0.53	307	98%	0.70	167	41	54.40	13.36
TSC_C07900	369	0.48	TTHA1292	369	0.55	125	97%	0.70	87	28	69.60	22.40
TSC_C24860	321	0.41	TTHA0061	321	0.48	108	97%	0.69	66	14	61.11	12.96
TSC_C13880	219	0.41	TTHA1034	219	0.48	72	99%	0.69	50	16	69.44	22.22
TSC_C00310	762	0.53	TTHA1011	795	0.60	277	92%	0.69	186	66	67.15	23.83
TSC_C19050	381	0.43	TTHA1461	381	0.50	127	98%	0.69	52	16	40.94	12.60
TSC_C23630	777	0.51	TTHA1934	777	0.58	264	97%	0.69	169	52	64.02	19.70
TSC_C18910	756	0.44	TTHA1449	756	0.51	254	98%	0.69	161	46	63.39	18.11
TSC_C06580	465	0.50	TTHA0141	444	0.57	158	94%	0.69	89	33	56.33	20.89
TSC_C06440	1086	0.51	TTHA0129	1086	0.58	369	98%	0.69	219	58	59.35	15.72
TSC_C00140	834	0.52	TTHA0192	840	0.59	293	94%	0.69	244	82	83.28	27.99
TSC_C09370	795	0.53	TTHA1050	792	0.59	264	99%	0.69	165	52	62.50	19.70
TSC_C20570	1107	0.47	TTHA1372	1083	0.54	375	97%	0.69	225	65	60.00	17.33
TSC_C22110	804	0.51	TTHA1658	771	0.58	288	90%	0.69	176	52	61.11	18.06
TSC_C01640	801	0.47	TTHA0331	804	0.53	274	97%	0.69	146	41	53.28	14.96
TSC_C24180	1089	0.51	TTHA1890	1065	0.58	365	98%	0.69	191	65	52.33	17.81
TSC_C14770	1029	0.52	TTHA0747	1029	0.58	348	98%	0.68	201	49	57.76	14.08
TSC_C05560	954	0.48	TTHA1429	954	0.54	320	99%	0.68	192	56	60.00	17.50

TSC_C06420	576	0.42	TTHA0127	585	0.49	194	99%	0.68	92	19	47.42	9.79
TSC_C01660	552	0.47	TTHA0329	558	0.53	195	94%	0.68	134	39	68.72	20.00
TSC_C04080	726	0.50	TTHA1825	714	0.57	244	98%	0.68	154	44	63.11	18.03
TSC_C19470	1545	0.51	TTHA1542	1524	0.58	530	96%	0.68	325	90	61.32	16.98
TSC_C22120	1383	0.55	TTHA1659	1362	0.62	482	94%	0.68	349	108	72.41	22.41
TSC_C23450	1689	0.48	TTHA1852	1692	0.55	567	99%	0.68	242	76	42.68	13.40
TSC_C05350	1122	0.48	TTHA0497	1098	0.55	384	96%	0.68	243	62	63.28	16.15
TSC_C17100	279	0.48	TTHA0845	279	0.54	93	98%	0.67	40	13	43.01	13.98
TSC_C20370	447	0.36	TTHA1479	447	0.42	148	99%	0.67	70	19	47.30	12.84
TSC_C22930	1788	0.49	TTHA0373	1809	0.56	606	99%	0.67	329	81	54.29	13.37
TSC_C21990	825	0.47	TTHA1568	819	0.54	277	98%	0.67	159	45	57.40	16.25
TSC_C17070	438	0.45	TTHA0849	444	0.52	147	99%	0.67	67	16	45.58	10.88
TSC_C17820	855	0.46	TTHA1069	855	0.53	289	98%	0.67	178	49	61.59	16.96
TSC_C23400	1713	0.46	TTHA1647	1656	0.53	581	96%	0.67	341	102	58.69	17.56
TSC_C09970	1395	0.47	TTHA0634	1395	0.54	465	100%	0.67	238	49	51.18	10.54
TSC_C17990	954	0.53	TTHA1100	930	0.60	326	96%	0.67	195	59	59.82	18.10
TSC_C03360	600	0.45	TTHA0169	567	0.52	207	93%	0.67	110	41	53.14	19.81
TSC_C15280	438	0.43	TTHA1193	438	0.49	146	99%	0.67	103	27	70.55	18.49
TSC_C19850	639	0.44	TTHA0101	618	0.51	218	95%	0.67	129	39	59.17	17.89
TSC_C03380	561	0.50	TTHA0167	570	0.57	190	98%	0.67	93	26	48.95	13.68
TSC_C24190	1116	0.51	TTHA1891	1023	0.58	396	88%	0.66	197	57	49.75	14.39
TSC_C07320	891	0.55	TTHA0811	855	0.62	305	95%	0.66	185	52	60.66	17.05
TSC_C01900	318	0.40	TTHA1909	315	0.47	105	99%	0.66	40	10	38.10	9.52
TSC_C01280	591	0.45	TTHA1799	585	0.52	205	94%	0.66	117	41	57.07	20.00
TSC_C14880	765	0.51	TTHA0656	765	0.57	261	97%	0.66	159	47	60.92	18.01
TSC_C14860	894	0.43	TTHA0659	903	0.50	303	98%	0.66	149	50	49.17	16.50
TSC_C01870	222	0.33	TTHA1912	225	0.40	75	97%	0.66	47	14	62.67	18.67
TSC_C05540	438	0.43	TTHA1432	438	0.49	145	99%	0.66	79	23	54.48	15.86
TSC_C24540	921	0.52	TTHA0049	918	0.59	312	98%	0.66	171	45	54.81	14.42

TSC_C04960	984	0.58	TTHA1351	999	0.64	342	96%	0.66	195	57	57.02	16.67
TSC_C06200	927	0.52	TTHA0431	921	0.58	309	99%	0.66	148	37	47.90	11.97
TSC_C04290	1503	0.49	TTHA1807	1491	0.56	525	94%	0.66	280	75	53.33	14.29
TSC_C05380	1221	0.47	TTHA0500	1221	0.54	415	98%	0.66	253	77	60.96	18.55
TSC_C21290	1752	0.48	TTHA1596	1716	0.55	606	95%	0.66	360	117	59.41	19.31
TSC_C14490	1254	0.49	TTHA1082	1254	0.56	425	98%	0.66	247	64	58.12	15.06
TSC_C06150	936	0.52	TTHA0424	936	0.59	322	96%	0.66	183	44	56.83	13.66
TSC_C15370	675	0.43	TTHA0716	675	0.50	229	97%	0.66	140	37	61.14	16.16
TSC_C12080	1047	0.44	TTHA0512	1056	0.51	369	94%	0.66	308	96	83.47	26.02
TSC_C07440	2532	0.53	TTHA1249	2535	0.59	871	97%	0.66	500	163	57.41	18.71
TSC_C24110	1248	0.50	TTHA1884	1266	0.57	424	98%	0.66	180	37	42.45	8.73
TSC_C00100	414	0.43	TTHA0188	414	0.49	137	99%	0.65	57	9	41.61	6.57
TSC_C12720	1539	0.54	TTHA0988	1485	0.61	528	95%	0.65	343	77	64.96	14.58
TSC_C06620	894	0.52	TTHA0144	894	0.58	315	94%	0.65	216	73	68.57	23.17
TSC_C13980	696	0.52	TTHA0623	690	0.59	241	95%	0.65	164	44	68.05	18.26
TSC_C04610	993	0.50	TTHA0013	993	0.56	334	98%	0.65	236	76	70.66	22.75
TSC_C09480	693	0.49	TTHA0762	705	0.56	243	95%	0.65	173	51	71.19	20.99
TSC_C02640	576	0.51	TTHA0291	576	0.58	192	99%	0.65	105	38	54.69	19.79
TSC_C09380	579	0.47	TTHA1053	576	0.53	209	90%	0.65	98	30	46.89	14.35
TSC_C23130	756	0.45	TTHA1624	729	0.52	256	96%	0.65	129	41	50.39	16.02
TSC_C17020	630	0.47	TTHA0854	630	0.54	214	97%	0.65	145	44	67.76	20.56
TSC_C05420	1047	0.51	TTHA0504	1044	0.57	359	96%	0.65	228	72	63.51	20.06
TSC_C09010	750	0.44	TTHA0693	756	0.50	253	98%	0.65	132	40	52.17	15.81
TSC_C15690	660	0.40	TTHA0833	660	0.47	222	98%	0.65	159	37	71.62	16.67
TSC_C00680	489	0.50	TTHA1951	492	0.56	168	96%	0.64	103	34	61.31	20.24
TSC_C18600	441	0.45	TTHA1543	441	0.51	152	95%	0.64	105	32	69.08	21.05
TSC_C18770	675	0.50	TTHA0381	555	0.57	226	89%	0.64	112	35	49.56	15.49
TSC_C20240	636	0.45	TTHA1348	642	0.52	215	98%	0.64	113	30	52.56	13.95
TSC_C04820	210	0.44	TTHA0433	231	0.51	78	91%	0.64	36	11	46.15	14.10

TSC_C24320	915	0.55	TTHA0320	918	0.62	315	96%	0.64	205	67	65.08	21.27
TSC_C10460	504	0.51	TTHA0961	450	0.58	168	93%	0.64	96	26	57.14	15.48
TSC_C23900	717	0.50	TTHA0296	717	0.56	242	98%	0.64	127	33	52.48	13.64
TSC_C11920	1092	0.44	TTHA0980	1092	0.50	365	99%	0.64	176	30	48.22	8.22
TSC_C24840	1254	0.52	TTHA0318	1227	0.58	434	95%	0.64	265	86	61.06	19.82
TSC_C20210	774	0.49	TTHA1345	765	0.56	266	96%	0.64	194	66	72.93	24.81
TSC_C20840	1458	0.53	TTHA1393	1449	0.59	498	97%	0.64	291	67	58.43	13.45
TSC_C19940	318	0.40	TTHA0104	318	0.46	107	97%	0.64	53	19	49.53	17.76
TSC_C06560	213	0.38	TTHA0139	234	0.45	78	93%	0.64	30	7	38.46	8.97
TSC_C15610	1035	0.49	TTHA1068	1035	0.55	349	98%	0.64	198	61	56.73	17.48
TSC_C17790	1113	0.49	TTHA0763	1092	0.55	374	98%	0.64	209	59	55.88	15.78
TSC_C04590	597	0.54	TTHA0014	597	0.60	205	96%	0.64	109	32	53.17	15.61
TSC_C22000	2784	0.54	TTHA1569	2751	0.60	939	98%	0.64	487	129	51.86	13.74
TSC_C22770	1035	0.53	TTHA0400	1035	0.59	344	100%	0.63	145	37	42.15	10.76
TSC_C02130	402	0.43	TTHA1794	381	0.49	136	94%	0.63	64	22	47.06	16.18
TSC_C22250	483	0.50	TTHA1673	453	0.56	160	96%	0.63	59	12	36.88	7.50
TSC_C19990	1956	0.47	TTHA0108	1956	0.53	657	99%	0.63	323	74	49.16	11.26
TSC_C07120	624	0.44	TTHA0790	627	0.50	220	94%	0.63	141	51	64.09	23.18
TSC_C18880	681	0.45	TTHA0374	651	0.52	229	96%	0.63	124	35	54.15	15.28
TSC_C15210	684	0.50	TTHA1186	663	0.56	229	97%	0.63	137	38	59.83	16.59
TSC_C22570	939	0.49	TTHA0217	939	0.56	313	99%	0.63	173	36	55.27	11.50
TSC_C22650	1389	0.48	TTHA1844	1389	0.54	462	100%	0.63	191	35	41.34	7.58
TSC_C03050	258	0.41	TTHA0032	243	0.47	92	87%	0.63	76	24	82.61	26.09
TSC_C09180	426	0.48	TTHA0702	474	0.54	158	93%	0.63	74	16	46.84	10.13
TSC_C09810	456	0.45	TTHA1121	456	0.51	152	99%	0.63	89	28	58.55	18.42
TSC_C20850	879	0.52	TTHA1394	879	0.58	297	98%	0.62	164	36	55.22	12.12
TSC_C12220	1533	0.55	TTHA0983	1482	0.61	534	93%	0.62	331	88	61.99	16.48
TSC_C11780	765	0.46	TTHA0671	762	0.52	259	97%	0.62	170	59	65.64	22.78
TSC_C08770	321	0.46	TTHA1274	315	0.53	107	97%	0.62	58	18	54.21	16.82

TSC_C23640	726	0.45	TTHA1933	732	0.52	247	98%	0.62	130	43	52.63	17.41
TSC_C09320	417	0.48	TTHA1045	420	0.54	139	99%	0.62	65	12	46.76	8.63
TSC_C19860	495	0.51	TTHA0102	504	0.58	177	92%	0.62	105	31	59.32	17.51
TSC_C03130	552	0.47	TTHA1553	567	0.53	203	90%	0.62	163	63	80.30	31.03
TSC_C01180	258	0.44	TTHA0253	258	0.50	87	97%	0.62	72	26	82.76	29.89
TSC_C01340	1560	0.48	TTHA0199	1548	0.54	550	93%	0.62	381	122	69.27	22.18
TSC_C01610	333	0.45	TTHA0334	333	0.51	112	97%	0.62	49	17	43.75	15.18
TSC_C09600	2439	0.49	TTHA1141	2430	0.55	831	97%	0.62	502	131	60.41	15.76
TSC_C18420	474	0.43	TTHA0776	417	0.50	160	91%	0.61	109	34	68.13	21.25
TSC_C03240	1617	0.52	TTHA1766	1410	0.59	544	92%	0.61	304	98	55.88	18.01
TSC_C07610	540	0.44	TTHA1231	540	0.50	180	99%	0.61	118	39	65.56	21.67
TSC_C09470	1296	0.51	TTHA0925	1296	0.57	446	96%	0.61	303	100	67.94	22.42
TSC_C09890	849	0.51	TTHA1112	852	0.57	294	96%	0.61	196	66	66.67	22.45
TSC_C13660	1146	0.54	TTHA0901	1155	0.60	405	94%	0.61	299	105	73.83	25.93
TSC_C03530	240	0.37	TTHA1913	207	0.43	87	81%	0.61	64	24	73.56	27.59
TSC_C19500	1587	0.45	TTHA0481	1590	0.51	537	98%	0.61	289	88	53.82	16.39
TSC_C20530	807	0.47	TTHA1368	807	0.53	271	99%	0.61	149	40	54.98	14.76
TSC_C20390	474	0.48	TTHA1476	474	0.54	157	99%	0.61	75	20	47.77	12.74
TSC_C22960	543	0.48	TTHA0370	543	0.54	184	97%	0.61	88	17	47.83	9.24
TSC_C18000	840	0.47	TTHA1101	861	0.53	287	98%	0.61	124	33	43.21	11.50
TSC_C19880	855	0.52	TTHA0103	855	0.58	285	99%	0.61	112	23	39.30	8.07
TSC_C19490	210	0.44	TTHA1309	207	0.50	69	98%	0.60	41	14	59.42	20.29
TSC_C05450	705	0.51	TTHA0507	825	0.58	287	87%	0.60	155	48	54.01	16.72
TSC_C20030	1566	0.52	TTHA0109	1554	0.58	538	96%	0.60	300	91	55.76	16.91
TSC_C14190	1059	0.49	TTHA0582	1059	0.55	359	98%	0.60	202	49	56.27	13.65
TSC_C01300	1296	0.50	TTHA1797	1305	0.56	435	99%	0.60	214	49	49.20	11.26
TSC_C22820	1002	0.47	TTHA0395	990	0.53	358	92%	0.60	257	92	71.79	25.70
TSC_C23020	1140	0.47	TTHA0072	1119	0.53	389	96%	0.60	250	77	64.27	19.79
TSC_C11870	1614	0.46	TTHA0604	1626	0.52	547	98%	0.60	255	52	46.62	9.51

TSC_C17940	801	0.41	TTHA1096	813	0.47	277	96%	0.60	174	41	62.82	14.80
TSC_C01150	1398	0.51	TTHA0256	1578	0.57	537	91%	0.59	266	81	49.53	15.08
TSC_C10320	843	0.45	TTHA0951	843	0.51	284	98%	0.59	170	51	59.86	17.96
TSC_C21420	450	0.50	TTHA1609	537	0.56	183	88%	0.59	101	30	55.19	16.39
TSC_C19970	669	0.44	TTHA0106	663	0.50	228	96%	0.59	122	27	53.51	11.84
TSC_C20640	519	0.45	TTHA1377	492	0.51	180	92%	0.59	121	25	67.22	13.89
TSC_C13690	327	0.46	TTHA0898	327	0.52	108	99%	0.59	70	21	64.81	19.44
TSC_C06520	483	0.47	TTHA0135	465	0.53	167	93%	0.59	97	22	58.08	13.17
TSC_C19130	786	0.49	TTHA0407	786	0.55	264	98%	0.59	117	23	44.32	8.71
TSC_C10590	2589	0.53	TTHA0663	2598	0.59	876	98%	0.59	439	116	50.11	13.24
TSC_C08820	753	0.53	TTHA0728	747	0.58	257	96%	0.59	162	57	63.04	22.18
TSC_C13670	486	0.42	TTHA0900	498	0.48	181	88%	0.59	124	50	68.51	27.62
TSC_C14570	501	0.52	TTHA1090	501	0.58	167	99%	0.58	76	18	45.51	10.78
TSC_C13050	423	0.50	TTHA1147	462	0.56	168	85%	0.58	117	42	69.64	25.00
TSC_C14130	1551	0.47	TTHA0608	1551	0.53	523	98%	0.58	321	95	61.38	18.16
TSC_C10140	393	0.50	TTHA1023	336	0.56	138	85%	0.58	87	34	63.04	24.64
TSC_C19170	1854	0.47	TTHA0403	1857	0.52	621	99%	0.58	265	58	42.67	9.34
TSC_C11760	669	0.39	TTHA0669	669	0.45	224	99%	0.58	133	38	59.38	16.96
TSC_C04770	2847	0.42	TTHA1016	2745	0.48	1141	77%	0.58	866	322	75.90	28.22
TSC_C00230	1011	0.49	TTHA0004	981	0.55	353	93%	0.58	218	65	61.76	18.41
TSC_C14150	849	0.43	TTHA0605	858	0.49	288	98%	0.58	148	33	51.39	11.46
TSC_C05620	915	0.49	TTHA0436	915	0.55	304	100%	0.58	128	30	42.11	9.87
TSC_C24340	819	0.56	TTHA0322	810	0.61	293	91%	0.58	187	61	63.82	20.82
TSC_C13040	942	0.49	TTHA1146	951	0.54	319	98%	0.58	165	37	51.72	11.60
TSC_C02140	1485	0.53	TTHA0201	1485	0.59	502	98%	0.58	212	55	42.23	10.96
TSC_C21070	1662	0.47	TTHA1417	1722	0.52	576	98%	0.58	287	82	49.83	14.24
TSC_C22840	402	0.51	TTHA0392	393	0.56	136	96%	0.58	64	13	47.06	9.56
TSC_C18080	2604	0.49	TTHA0826	2604	0.55	906	95%	0.58	470	142	51.88	15.67
TSC_C07150	594	0.47	TTHA0793	588	0.53	202	96%	0.58	112	26	55.45	12.87

TSC_C17880	684	0.49	TTHA1299	684	0.55	231	98%	0.57	118	25	51.08	10.82
TSC_C23500	276	0.50	TTHA1857	258	0.56	96	90%	0.57	58	25	60.42	26.04
TSC_C24300	636	0.46	TTHA1921	597	0.52	229	88%	0.57	185	63	80.79	27.51
TSC_C21120	342	0.46	TTHA1422	462	0.52	155	83%	0.57	62	20	40.00	12.90
TSC_C21710	639	0.54	TTHA1507	681	0.59	234	93%	0.57	132	34	56.41	14.53
TSC_C14360	909	0.52	TTHA0572	909	0.58	304	99%	0.57	169	47	55.59	15.46
TSC_C01560	483	0.45	TTHA0363	486	0.51	165	97%	0.57	75	23	45.45	13.94
TSC_C07570	393	0.49	TTHA1235	387	0.55	136	94%	0.57	100	38	73.53	27.94
TSC_C24900	660	0.52	TTHA0065	660	0.58	224	97%	0.57	160	44	71.43	19.64
TSC_C07590	735	0.50	TTHA1233	735	0.56	246	99%	0.57	144	42	58.54	17.07
TSC_C22680	663	0.53	TTHA1580	663	0.59	223	98%	0.57	94	20	42.15	8.97
TSC_C15240	789	0.55	TTHA1189	789	0.61	270	97%	0.57	175	49	64.81	18.15
TSC_C11830	963	0.52	TTHA0677	939	0.57	328	96%	0.57	224	67	68.29	20.43
TSC_C05470	1314	0.49	TTHA0509	1299	0.54	441	98%	0.57	220	71	49.89	16.10
TSC_C13410	513	0.52	TTHA0928	558	0.57	187	94%	0.57	83	24	44.39	12.83
TSC_C24910	1062	0.48	TTHA1958	1053	0.54	356	98%	0.57	167	44	46.91	12.36
TSC_C09720	1314	0.49	TTHA1130	1326	0.55	444	99%	0.57	248	61	55.86	13.74
TSC_C04500	2088	0.53	TTHA0023	2094	0.59	709	98%	0.56	372	111	52.47	15.66
TSC_C23540	624	0.49	TTHA1862	630	0.54	213	97%	0.56	162	46	76.06	21.60
TSC_C13320	279	0.43	TTHA0876	279	0.48	94	97%	0.56	54	15	57.45	15.96
TSC_C18300	546	0.44	TTHA0783	546	0.50	181	99%	0.56	77	18	42.54	9.94
TSC_C21900	477	0.44	TTHA1559	495	0.50	165	97%	0.56	78	25	47.27	15.15
TSC_C12120	915	0.49	TTHA0974	942	0.55	318	97%	0.56	151	32	47.48	10.06
TSC_C20750	1368	0.52	TTHA1387	1371	0.58	461	99%	0.56	258	67	55.97	14.53
TSC_C13900	1305	0.49	TTHA1036	1305	0.55	434	100%	0.56	208	50	47.93	11.52
TSC_C22780	411	0.49	TTHA0399	417	0.54	140	97%	0.56	77	22	55.00	15.71
TSC_C06610	663	0.42	TTHA0143	663	0.48	220	100%	0.56	101	29	45.91	13.18
TSC_C23880	516	0.49	TTHA0294	516	0.55	171	99%	0.55	108	24	63.16	14.04
TSC_C19450	609	0.48	TTHA1530	630	0.54	222	92%	0.55	159	50	71.62	22.52

TSC_C21320	318	0.42	TTHA1599	318	0.47	105	99%	0.55	48	13	45.71	12.38
TSC_C00920	1014	0.45	TTHA0311	1014	0.50	351	96%	0.55	191	64	54.42	18.23
TSC_C24160	879	0.52	TTHA1888	879	0.58	294	99%	0.55	157	49	53.40	16.67
TSC_C23270	717	0.51	TTHA1638	717	0.57	242	98%	0.55	148	44	61.16	18.18
TSC_C19080	1164	0.52	TTHA0456	1125	0.58	392	97%	0.55	199	48	50.77	12.24
TSC_C14530	798	0.48	TTHA1086	798	0.53	267	99%	0.55	121	24	45.32	8.99
TSC_C13170	1446	0.44	TTHA0893	1443	0.50	486	99%	0.55	204	46	41.98	9.47
TSC_C06540	375	0.46	TTHA0137	375	0.51	124	99%	0.55	67	18	54.03	14.52
TSC_C18110	612	0.44	TTHA0823	636	0.49	213	97%	0.55	105	28	49.30	13.15
TSC_C07310	897	0.51	TTHA0810	900	0.56	302	98%	0.55	167	50	55.30	16.56
TSC_C12140	588	0.47	TTHA0973	612	0.52	204	97%	0.55	111	29	54.41	14.22
TSC_C20250	294	0.37	TTHA1349	294	0.42	97	99%	0.55	34	7	35.05	7.22
TSC_C03070	258	0.45	TTHA1782	258	0.51	85	99%	0.55	31	4	36.47	4.71
TSC_C08660	1899	0.52	TTHA1285	1890	0.58	645	98%	0.55	344	90	53.33	13.95
TSC_C13340	1110	0.49	TTHA0874	1110	0.54	370	99%	0.55	158	41	42.70	11.08
TSC_C21230	483	0.56	TTHA1590	426	0.62	180	80%	0.55	86	29	47.78	16.11
TSC_C08760	972	0.53	TTHA1275	972	0.58	325	99%	0.55	186	53	57.23	16.31
TSC_C00340	468	0.47	TTHA0346	450	0.53	155	97%	0.55	50	15	32.26	9.68
TSC_C05430	1404	0.50	TTHA0505	1383	0.56	469	99%	0.55	233	53	49.68	11.30
TSC_C09460	606	0.46	TTHA0926	612	0.51	208	97%	0.54	107	35	51.44	16.83
TSC_C02860	636	0.46	TTHA1107	630	0.52	219	95%	0.54	129	45	58.90	20.55
TSC_C01120	357	0.39	TTHA1201	342	0.44	125	91%	0.54	122	50	97.60	40.00
TSC_C18320	1281	0.48	TTHA0781	1281	0.53	429	99%	0.54	261	59	60.84	13.75
TSC_C09360	1269	0.50	TTHA1049	1302	0.56	436	98%	0.54	237	54	54.36	12.39
TSC_C22100	636	0.45	TTHA1657	636	0.50	213	99%	0.54	98	13	46.01	6.10
TSC_C20720	555	0.45	TTHA1385	555	0.50	185	99%	0.54	98	31	52.97	16.76
TSC_C08000	1215	0.53	TTHA1265	1212	0.58	410	98%	0.54	208	51	50.73	12.44
TSC_C09420	399	0.48	TTHA0908	399	0.54	133	98%	0.54	67	13	50.38	9.77
TSC_C20510	1383	0.47	TTHA1366	1389	0.53	464	99%	0.54	192	50	41.38	10.78

TSC_C13280	1212	0.43	TTHA0882	1212	0.48	405	99%	0.54	215	40	53.09	9.88
TSC_C15200	1125	0.48	TTHA1185	1107	0.53	377	98%	0.54	168	36	44.56	9.55
TSC_C22800	1044	0.49	TTHA0397	1044	0.55	353	98%	0.54	189	58	53.54	16.43
TSC_C03930	2763	0.52	TTHA1837	2757	0.58	940	98%	0.54	538	157	57.23	16.70
TSC_C05030	1011	0.42	TTHA1337	921	0.47	355	89%	0.54	282	99	79.44	27.89
TSC_C21470	528	0.45	TTHA1613	528	0.50	178	98%	0.54	83	21	46.63	11.80
TSC_C18650	1779	0.49	TTHA1548	1773	0.54	596	99%	0.54	276	65	46.31	10.91
TSC_C00190	321	0.46	TTHA1788	330	0.52	114	93%	0.54	77	27	67.54	23.68
TSC_C21190	432	0.39	TTHA1465	423	0.44	143	98%	0.54	55	16	38.46	11.19
TSC_C01320	429	0.54	TTHA1795	417	0.59	144	96%	0.54	88	29	61.11	20.14
TSC_C20450	1557	0.54	TTHA1471	1608	0.60	555	94%	0.53	299	84	53.87	15.14
TSC_C05510	708	0.52	TTHA1435	708	0.57	238	98%	0.53	147	44	61.76	18.49
TSC_C18800	1194	0.48	TTHA0378	1137	0.53	403	96%	0.53	163	31	40.45	7.69
TSC_C08920	1932	0.48	TTHA0738	1926	0.54	678	94%	0.53	516	176	76.11	25.96
TSC_C11010	1119	0.46	TTHA1137	1149	0.51	459	78%	0.53	329	126	71.68	27.45
TSC_C05130	705	0.41	TTHA1326	693	0.47	238	97%	0.53	161	42	67.65	17.65
TSC_C04090	1080	0.48	TTHA1824	1095	0.53	367	98%	0.53	179	50	48.77	13.62
TSC_C22860	942	0.52	TTHA0389	948	0.57	319	98%	0.53	155	34	48.59	10.66
TSC_C09490	411	0.46	TTHA0761	411	0.52	138	98%	0.53	105	28	76.09	20.29
TSC_C08630	1092	0.47	TTHA1287	1029	0.52	367	96%	0.53	208	64	56.68	17.44
TSC_C06990	885	0.50	TTHA0527	885	0.55	298	98%	0.53	161	43	54.03	14.43
TSC_C17050	1200	0.51	TTHA0851	1197	0.56	407	98%	0.53	251	63	61.67	15.48
TSC_C09300	363	0.43	TTHA1043	363	0.49	123	97%	0.53	69	21	56.10	17.07
TSC_C24100	684	0.54	TTHA1883	696	0.59	240	95%	0.53	157	41	65.42	17.08
TSC_C00010	1344	0.42	TTHA1973	1311	0.48	447	99%	0.53	183	46	40.94	10.29
TSC_C08230	735	0.49	TTHA0546	708	0.54	247	96%	0.53	145	39	58.70	15.79
TSC_C14430	861	0.48	TTHA1076	858	0.54	291	98%	0.52	189	59	64.95	20.27
TSC_C08250	429	0.51	TTHA0548	429	0.56	146	97%	0.52	93	31	63.70	21.23
TSC_C24610	351	0.44	TTHA0057	351	0.49	116	99%	0.52	45	9	38.79	7.76

TSC_C21330	585	0.52	TTHA1600	585	0.58	195	99%	0.52	109	34	55.90	17.44
TSC_C20060	654	0.47	TTHA0112	663	0.52	220	99%	0.52	122	28	55.45	12.73
TSC_C07750	309	0.48	TTHA0560	309	0.53	103	98%	0.52	57	16	55.34	15.53
TSC_C09860	543	0.48	TTHA1116	537	0.53	185	96%	0.52	119	34	64.32	18.38
TSC_C15080	1305	0.49	TTHA1175	1305	0.55	435	100%	0.52	217	51	49.89	11.72
TSC_C14950	315	0.46	TTHA1166	318	0.51	112	92%	0.52	90	31	80.36	27.68
TSC_C09830	456	0.44	TTHA1119	456	0.50	157	95%	0.52	90	28	57.32	17.83
TSC_C22600	537	0.40	TTHA0220	546	0.46	181	99%	0.52	57	11	31.49	6.08
TSC_C08070	1014	0.49	TTHA0540	1011	0.54	343	98%	0.52	201	57	58.60	16.62
TSC_C05460	699	0.45	TTHA0508	702	0.50	239	97%	0.52	55	12	23.01	5.02
TSC_C21090	783	0.46	TTHA1419	777	0.51	265	97%	0.52	125	28	47.17	10.57
TSC_C08740	300	0.44	TTHA1277	300	0.49	99	99%	0.52	41	9	41.41	9.09
TSC_C15580	672	0.45	TTHA1066	672	0.50	227	98%	0.52	98	21	43.17	9.25
TSC_C02230	813	0.49	TTHA0244	792	0.54	273	97%	0.52	134	39	49.08	14.29
TSC_C01060	1173	0.50	TTHA0263	1173	0.55	401	97%	0.52	212	53	52.87	13.22
TSC_C23290	1893	0.46	TTHA1640	1887	0.52	637	99%	0.52	256	66	40.19	10.36
TSC_C02610	681	0.45	TTHA1007	687	0.50	228	99%	0.52	126	31	55.26	13.60
TSC_C24020	693	0.54	TTHA1879	693	0.59	233	98%	0.52	128	31	54.94	13.30
TSC_C18640	963	0.51	TTHA1547	963	0.56	329	97%	0.52	216	70	65.65	21.28
TSC_C20560	1065	0.53	TTHA1371	1029	0.58	371	93%	0.51	250	82	67.39	22.10
TSC_C20780	366	0.39	TTHA1389	354	0.44	124	95%	0.51	79	24	63.71	19.35
TSC_C03370	495	0.47	TTHA0168	495	0.52	164	99%	0.51	56	7	34.15	4.27
TSC_C14540	585	0.50	TTHA1087	585	0.55	212	90%	0.51	160	55	75.47	25.94
TSC_C22560	1110	0.49	TTHA0216	1110	0.54	371	99%	0.51	144	20	38.81	5.39
TSC_C19610	1434	0.50	TTHA0495	1428	0.55	496	96%	0.51	299	88	60.28	17.74
TSC_C21010	1284	0.48	TTHA1411	1293	0.53	433	99%	0.51	183	41	42.26	9.47
TSC_C23910	1071	0.54	TTHA0297	1071	0.59	359	99%	0.51	180	54	50.14	15.04
TSC_C09440	2241	0.50	TTHA0910	2235	0.55	753	99%	0.51	388	87	51.53	11.55
TSC_C15060	1473	0.49	TTHA1173	1473	0.54	512	95%	0.51	290	94	56.64	18.36

TSC_C08910	855	0.48	TTHA0737	852	0.53	285	99%	0.51	159	43	55.79	15.09
TSC_C23700	477	0.48	TTHA1699	477	0.53	160	98%	0.51	112	32	70.00	20.00
TSC_C24660	1050	0.45	TTHA1937	1068	0.50	357	98%	0.51	155	48	43.42	13.45
TSC_C15630	816	0.51	TTHA0834	813	0.56	276	98%	0.51	146	40	52.90	14.49
TSC_C08750	600	0.54	TTHA1276	567	0.59	207	93%	0.51	120	41	57.97	19.81
TSC_C05960	540	0.47	TTHA0093	531	0.52	187	94%	0.51	85	22	45.45	11.76
TSC_C23690	1116	0.51	TTHA1698	1125	0.56	378	98%	0.51	194	45	51.32	11.90
TSC_C23190	1053	0.54	TTHA1630	1053	0.59	356	98%	0.51	178	52	50.00	14.61
TSC_C18410	792	0.48	TTHA0775	783	0.53	269	97%	0.51	183	53	68.03	19.70
TSC_C19960	924	0.45	TTHA1773	918	0.50	323	94%	0.51	228	98	70.59	30.34
TSC_C21340	411	0.51	TTHA1601	384	0.56	146	88%	0.51	96	33	65.75	22.60
TSC_C19400	1683	0.47	TTHA1557	1662	0.52	560	99%	0.50	212	41	37.86	7.32
TSC_C17080	441	0.53	TTHA0848	429	0.58	154	92%	0.50	101	38	65.58	24.68
TSC_C20730	1035	0.50	TTHA1386	1047	0.55	356	97%	0.50	211	62	59.27	17.42
TSC_C13650	435	0.38	TTHA0902	435	0.43	146	98%	0.50	63	11	43.15	7.53
TSC_C05010	1053	0.51	TTHA1339	1044	0.56	361	96%	0.50	253	91	70.08	25.21
TSC_C19160	1353	0.54	TTHA0404	1374	0.59	465	97%	0.50	263	82	56.56	17.63
TSC_C20050	1155	0.47	TTHA0111	1134	0.52	396	96%	0.50	277	75	69.95	18.94
TSC_C07180	1368	0.53	TTHA0797	1356	0.59	465	97%	0.50	271	84	58.28	18.06
TSC_C01680	879	0.49	TTHA0327	891	0.54	298	98%	0.50	115	29	38.59	9.73
TSC_C20500	711	0.49	TTHA1365	711	0.54	239	98%	0.50	116	37	48.54	15.48
TSC_C02280	675	0.48	TTHA0240	654	0.53	226	97%	0.50	99	24	43.81	10.62
TSC_C07890	399	0.47	TTHA1293	429	0.52	143	95%	0.50	76	20	53.15	13.99
TSC_C10250	387	0.41	TTHA0864	387	0.46	129	98%	0.50	55	17	42.64	13.18
TSC_C11800	621	0.48	TTHA0674	621	0.53	207	99%	0.50	105	22	50.72	10.63
TSC_C04040	1014	0.50	TTHA1828	984	0.55	343	96%	0.50	150	41	43.73	11.95
TSC_C02580	1263	0.49	TTHA1004	1257	0.54	453	92%	0.50	388	136	85.65	30.02
TSC_C09980	384	0.46	TTHA0633	384	0.51	128	98%	0.49	88	19	68.75	14.84
TSC_C14910	606	0.45	TTHA0653	597	0.50	207	96%	0.49	140	49	67.63	23.67

TSC_C01800	522	0.49	TTHA0209	522	0.54	174	99%	0.49	67	17	38.51	9.77
TSC_C06290	402	0.45	TTHA0519	378	0.50	138	92%	0.49	50	9	36.23	6.52
TSC_C13950	2586	0.53	TTHA0626	2577	0.58	872	98%	0.49	492	116	56.42	13.30
TSC_C20170	774	0.41	TTHA0459	774	0.46	268	95%	0.49	141	43	52.61	16.04
TSC_C01130	294	0.43	TTHA1512	273	0.47	103	89%	0.49	82	27	79.61	26.21
TSC_C18590	1056	0.49	TTHA1531	1071	0.54	364	97%	0.49	171	48	46.98	13.19
TSC_C00090	933	0.51	TTHA0187	936	0.56	324	95%	0.49	198	52	61.11	16.05
TSC_C20910	981	0.52	TTHA1401	981	0.57	328	99%	0.49	202	54	61.59	16.46
TSC_C21890	612	0.48	TTHA1558	612	0.52	207	98%	0.49	99	24	47.83	11.59
TSC_C24150	1203	0.50	TTHA1887	1224	0.55	411	98%	0.49	208	55	50.61	13.38
TSC_C23620	1218	0.49	TTHA1935	1197	0.54	409	98%	0.49	211	75	51.59	18.34
TSC_C07330	1248	0.53	TTHA0812	1254	0.58	420	99%	0.49	201	48	47.86	11.43
TSC_C24310	1722	0.42	TTHA1922	1518	0.47	626	84%	0.49	417	128	66.61	20.45
TSC_C06140	717	0.53	TTHA0423	717	0.58	242	98%	0.49	161	50	66.53	20.66
TSC_C09430	1251	0.49	TTHA0909	1245	0.54	416	100%	0.49	169	41	40.63	9.86
TSC_C01850	1131	0.45	TTHA1914	1131	0.50	376	100%	0.48	129	25	34.31	6.65
TSC_C10610	1302	0.44	TTHA0768	1314	0.49	508	83%	0.48	405	159	79.72	31.30
TSC_C24040	867	0.53	TTHA1881	864	0.58	290	99%	0.48	138	34	47.59	11.72
TSC_C24970	531	0.45	TTHA1965	528	0.50	178	98%	0.48	53	4	29.78	2.25
TSC_C15480	1317	0.47	TTHA0708	1317	0.52	442	99%	0.48	165	35	37.33	7.92
TSC_C00130	819	0.50	TTHA0191	885	0.55	295	95%	0.48	135	42	45.76	14.24
TSC_C14120	774	0.49	TTHA0609	768	0.54	265	96%	0.48	138	40	52.08	15.09
TSC_C07280	1065	0.50	TTHA1939	1065	0.55	356	99%	0.48	132	26	37.08	7.30
TSC_C23750	699	0.46	TTHA1705	684	0.50	236	97%	0.48	143	46	60.59	19.49
TSC_C22580	786	0.49	TTHA0218	795	0.54	267	98%	0.48	154	40	57.68	14.98
TSC_C03520	1089	0.47	TTHA1757	1086	0.52	362	100%	0.48	149	40	41.16	11.05
TSC_C00390	672	0.50	TTHA0340	693	0.55	235	96%	0.48	127	47	54.04	20.00
TSC_C21250	1305	0.56	TTHA1592	1305	0.61	451	96%	0.48	293	86	64.97	19.07
TSC_C03670	969	0.55	TTHA1740	969	0.60	331	97%	0.48	169	45	51.06	13.60

TSC_C06240	957	0.48	TTHA0515	957	0.53	322	98%	0.48	208	55	64.60	17.08
TSC_C15410	1746	0.49	TTHA0711	1743	0.54	582	100%	0.48	277	74	47.59	12.71
TSC_C13460	1299	0.50	TTHA0931	1299	0.55	439	98%	0.48	281	67	64.01	15.26
TSC_C01530	1410	0.46	TTHA0366	1410	0.51	470	100%	0.48	220	50	46.81	10.64
TSC_C00040	1425	0.50	TTHA0003	1425	0.55	481	98%	0.48	210	55	43.66	11.43
TSC_C21110	453	0.44	TTHA1421	459	0.49	152	99%	0.48	79	26	51.97	17.11
TSC_C21220	2211	0.56	TTHA0999	2211	0.61	771	95%	0.48	501	155	64.98	20.10
TSC_C14550	1245	0.49	TTHA1088	1236	0.54	419	98%	0.48	211	49	50.36	11.69
TSC_C08010	1221	0.50	TTHA1264	1221	0.55	411	99%	0.48	208	58	50.61	14.11
TSC_C23120	636	0.44	TTHA1623	624	0.49	216	96%	0.47	126	47	58.33	21.76
TSC_C18220	900	0.51	TTHA0788	879	0.56	302	97%	0.47	173	54	57.28	17.88
TSC_C04060	492	0.45	TTHA1826	498	0.50	168	97%	0.47	94	27	55.95	16.07
TSC_C17540	654	0.51	TTHA1437	651	0.56	250	84%	0.47	216	97	86.40	38.80
TSC_C19840	2859	0.48	TTHA1440	2859	0.53	953	100%	0.47	334	63	35.05	6.61
TSC_C18090	420	0.47	TTHA0825	402	0.52	139	97%	0.47	47	12	33.81	8.63
TSC_C24560	855	0.53	TTHA0051	855	0.57	289	98%	0.47	152	42	52.60	14.53
TSC_C08900	669	0.50	TTHA0736	675	0.55	232	96%	0.47	150	38	64.66	16.38
TSC_C15330	1047	0.48	TTHA1197	1038	0.52	353	98%	0.47	202	51	57.22	14.45
TSC_C20330	1185	0.50	TTHA1485	1185	0.55	404	97%	0.47	202	69	50.00	17.08
TSC_C24470	1098	0.47	TTHA0309	1098	0.51	373	98%	0.47	219	71	58.71	19.03
TSC_C03790	621	0.49	TTHA1729	663	0.53	236	89%	0.47	169	60	71.61	25.42
TSC_C00650	471	0.49	TTHA1947	468	0.54	162	95%	0.47	82	23	50.62	14.20
TSC_C18660	1779	0.46	TTHA0098	1779	0.50	596	99%	0.47	296	77	49.66	12.92
TSC_C18020	1023	0.52	TTHA1103	1023	0.56	346	98%	0.47	205	59	59.25	17.05
TSC_C09250	420	0.46	TTHA1038	429	0.51	142	98%	0.47	47	5	33.10	3.52
TSC_C05110	1659	0.48	TTHA1328	1659	0.52	562	98%	0.47	225	51	40.04	9.07
TSC_C22540	1566	0.52	TTHA0214	1569	0.57	534	97%	0.47	259	73	48.50	13.67
TSC_C21550	795	0.50	TTHA0077	795	0.54	270	97%	0.47	139	38	51.48	14.07
TSC_C18670	1050	0.55	TTHA0099	1047	0.60	354	98%	0.46	203	57	57.34	16.10

TSC_C07420	249	0.39	TTHA0820	258	0.44	90	91%	0.46	53	19	58.89	21.11
TSC_C22150	477	0.45	TTHA1662	480	0.49	168	93%	0.46	86	26	51.19	15.48
TSC_C20290	837	0.49	TTHA1489	843	0.53	282	99%	0.46	131	31	46.45	10.99
TSC_C08580	1239	0.53	TTHA0722	1239	0.58	412	100%	0.46	197	50	47.82	12.14
TSC_C13130	1263	0.49	TTHA1152	1227	0.53	425	97%	0.46	214	59	50.35	13.88
TSC_C13230	2943	0.50	TTHA0889	2937	0.55	987	99%	0.46	531	122	53.80	12.36
TSC_C17900	486	0.46	TTHA1091	486	0.50	162	99%	0.46	67	8	41.36	4.94
TSC_C24890	1050	0.49	TTHA0064	1059	0.54	363	96%	0.46	234	59	64.46	16.25
TSC_C15400	1257	0.51	TTHA0712	1266	0.55	422	99%	0.46	161	39	38.15	9.24
TSC_C08380	966	0.51	TTHA1215	981	0.56	330	98%	0.46	182	50	55.15	15.15
TSC_C00030	1269	0.49	TTHA0002	1269	0.53	423	100%	0.46	178	31	42.08	7.33
TSC_C03250	693	0.49	TTHA0179	699	0.54	235	98%	0.46	138	39	58.72	16.60
TSC_C09220	2076	0.57	TTHA1061	2121	0.62	725	96%	0.46	487	143	67.17	19.72
TSC_C07820	1659	0.53	TTHA0566	1512	0.58	561	93%	0.46	254	72	45.28	12.83
TSC_C19020	1152	0.50	TTHA1458	1146	0.54	385	99%	0.46	209	70	54.29	18.18
TSC_C15090	876	0.50	TTHA1176	858	0.55	305	94%	0.46	217	56	71.15	18.36
TSC_C22660	1551	0.49	TTHA1578	1551	0.53	522	99%	0.46	246	76	47.13	14.56
TSC_C19290	579	0.51	TTHA1588	552	0.56	195	95%	0.45	106	22	54.36	11.28
TSC_C00940	702	0.48	TTHA0313	720	0.52	246	95%	0.45	135	39	54.88	15.85
TSC_C13610	1170	0.52	TTHA0906	1173	0.57	396	98%	0.45	185	49	46.72	12.37
TSC_C19740	1008	0.45	TTHA0465	1008	0.50	339	99%	0.45	142	31	41.89	9.14
TSC_C20180	639	0.53	TTHA0458	627	0.57	215	97%	0.45	129	42	60.00	19.53
TSC_C13990	471	0.43	TTHA0622	468	0.47	156	99%	0.45	58	7	37.18	4.49
TSC_C03640	783	0.51	TTHA1743	774	0.56	265	97%	0.45	162	39	61.13	14.72
TSC_C13890	255	0.42	TTHA1035	267	0.46	88	97%	0.45	35	6	39.77	6.82
TSC_C14310	1701	0.49	TTHA0577	1701	0.53	576	98%	0.45	372	120	64.58	20.83
TSC_C01360	471	0.44	TTHA0197	480	0.48	162	97%	0.45	111	34	68.52	20.99
TSC_C18380	801	0.49	TTHA0772	810	0.54	276	96%	0.44	161	37	58.33	13.41
TSC_C06070	750	0.47	TTHA0082	744	0.51	256	96%	0.44	148	44	57.81	17.19

TSC_C00540	798	0.50	TTHA1941	795	0.55	270	98%	0.44	150	30	55.56	11.11
TSC_C10300	459	0.45	TTHA0949	459	0.49	154	98%	0.44	73	17	47.40	11.04
TSC_C23380	1329	0.42	TTHA1651	1329	0.46	452	98%	0.44	222	64	49.12	14.16
TSC_C14720	672	0.51	TTHA0754	660	0.56	226	97%	0.44	130	37	57.52	16.37
TSC_C12710	1203	0.48	TTHA0987	1206	0.53	402	99%	0.44	191	42	47.51	10.45
TSC_C13740	336	0.41	TTHA0467	429	0.46	149	82%	0.44	77	30	51.68	20.13
TSC_C09280	1476	0.44	TTHA1041	1479	0.48	494	99%	0.44	197	40	39.88	8.10
TSC_C23560	420	0.51	TTHA0316	444	0.56	152	93%	0.44	113	38	74.34	25.00
TSC_C07340	1914	0.53	TTHA0813	1875	0.57	677	92%	0.44	475	178	70.16	26.29
TSC_C12020	1593	0.52	TTHA0755	1575	0.56	536	98%	0.44	317	87	59.14	16.23
TSC_C01420	1818	0.45	TTHA0352	1827	0.50	618	98%	0.44	345	123	55.83	19.90
TSC_C24260	1038	0.47	TTHA1920	978	0.52	348	96%	0.44	131	28	37.64	8.05
TSC_C04110	1041	0.49	TTHA1822	1050	0.54	361	96%	0.44	235	75	65.10	20.78
TSC_C03660	1005	0.49	TTHA1741	1005	0.53	338	99%	0.44	233	70	68.93	20.71
TSC_C13450	1143	0.48	TTHA0759	1143	0.52	383	99%	0.43	160	36	41.78	9.40
TSC_C07600	534	0.45	TTHA1232	534	0.50	180	98%	0.43	96	30	53.33	16.67
TSC_C03770	243	0.37	TTHA1730	246	0.42	85	93%	0.43	53	18	62.35	21.18
TSC_C19760	1020	0.46	TTHA0464	1011	0.50	355	95%	0.43	219	70	61.69	19.72
TSC_C21630	1407	0.51	TTHA1503	1407	0.55	471	99%	0.43	244	69	51.80	14.65
TSC_C22590	711	0.51	TTHA0219	708	0.55	242	97%	0.43	139	35	57.44	14.46
TSC_C23010	717	0.50	TTHA0148	732	0.54	251	95%	0.43	156	50	62.15	19.92
TSC_C01080	870	0.46	TTHA0261	876	0.51	352	78%	0.43	238	100	67.61	28.41
TSC_C20920	432	0.47	TTHA1402	420	0.51	146	96%	0.43	65	11	44.52	7.53
TSC_C23110	693	0.50	TTHA1622	693	0.54	231	99%	0.43	118	25	51.08	10.82
TSC_C00240	615	0.48	TTHA0005	615	0.53	208	98%	0.43	135	45	64.90	21.63
TSC_C08610	1155	0.53	TTHA0725	1146	0.58	397	96%	0.43	247	67	62.22	16.88
TSC_C18510	1113	0.50	TTHA1536	750	0.54	376	78%	0.43	137	36	36.44	9.57
TSC_C15350	906	0.46	TTHA1199	906	0.51	303	99%	0.43	152	32	50.17	10.56
TSC_C22990	1128	0.51	TTHA0146	1128	0.55	385	97%	0.43	218	64	56.62	16.62

TSC_C14980	840	0.51	TTHA1161	828	0.56	285	97%	0.43	178	51	62.46	17.89
TSC_C01810	378	0.45	TTHA0210	378	0.50	126	98%	0.43	38	8	30.16	6.35
TSC_C24410	267	0.45	TTHA0302	210	0.50	89	86%	0.43	36	16	40.45	17.98
TSC_C12190	2010	0.52	TTHA0968	2007	0.56	676	99%	0.42	394	100	58.28	14.79
TSC_C08320	465	0.52	TTHA0554	471	0.57	162	95%	0.42	80	22	49.38	13.58
TSC_C08330	678	0.51	TTHA0555	678	0.56	241	92%	0.42	150	49	62.24	20.33
TSC_C24680	723	0.51	TTHA0160	702	0.56	251	93%	0.42	147	47	58.57	18.73
TSC_C05600	420	0.38	TTHA0435	426	0.43	147	94%	0.42	78	22	53.06	14.97
TSC_C09450	1116	0.51	TTHA0924	981	0.55	373	93%	0.42	159	41	42.63	10.99
TSC_C23770	747	0.48	TTHA1706	747	0.53	252	98%	0.42	120	37	47.62	14.68
TSC_C09730	873	0.48	TTHA1129	882	0.52	293	99%	0.42	164	35	55.97	11.95
TSC_C17980	1509	0.51	TTHA1099	1491	0.55	508	98%	0.42	254	61	50.00	12.01
TSC_C17010	1011	0.49	TTHA0855	1011	0.54	337	99%	0.42	177	42	52.52	12.46
TSC_C19410	864	0.49	TTHA1526	861	0.53	294	97%	0.42	170	46	57.82	15.65
TSC_C20900	555	0.48	TTHA1400	555	0.52	185	99%	0.42	88	21	47.57	11.35
TSC_C05580	2127	0.52	TTHA1427	2079	0.56	718	97%	0.42	332	93	46.24	12.95
TSC_C21380	1809	0.52	TTHA1605	1857	0.56	628	97%	0.42	321	85	51.11	13.54
TSC_C05280	1119	0.49	TTHA1314	1131	0.54	376	99%	0.42	163	38	43.35	10.11
TSC_C02490	2331	0.56	TTHA1716	2217	0.60	823	91%	0.42	529	198	64.28	24.06
TSC_C05260	1092	0.47	TTHA1316	1113	0.51	382	96%	0.42	234	69	61.26	18.06
TSC_C06430	624	0.47	TTHA0128	621	0.51	208	99%	0.42	122	29	58.65	13.94
TSC_C07830	267	0.50	TTHA0567	267	0.54	89	98%	0.42	31	8	34.83	8.99
TSC_C02440	1047	0.51	TTHA0224	1047	0.55	349	99%	0.42	182	56	52.15	16.05
TSC_C02560	660	0.53	TTHA1002	723	0.57	242	94%	0.42	107	32	44.21	13.22
TSC_C22980	441	0.42	TTHA0640	414	0.46	170	79%	0.42	134	45	78.82	26.47
TSC_C10540	423	0.47	TTHA0967	426	0.52	141	99%	0.42	65	17	46.10	12.06
TSC_C24750	564	0.47	TTHA1926	564	0.51	188	99%	0.41	68	15	36.17	7.98
TSC_C01260	495	0.48	TTHA0341	495	0.53	164	99%	0.41	82	18	50.00	10.98
TSC_C14060	1215	0.50	TTHA0614	1215	0.54	411	98%	0.41	204	69	49.64	16.79

TSC_C24940	969	0.52	TTHA1962	969	0.56	324	99%	0.41	128	27	39.51	8.33
TSC_C09960	303	0.48	TTHA0635	303	0.52	101	98%	0.41	69	22	68.32	21.78
TSC_C05150	2427	0.50	TTHA1324	2460	0.54	824	99%	0.41	431	87	52.31	10.56
TSC_C03060	306	0.44	TTHA1783	306	0.48	101	99%	0.41	51	14	50.50	13.86
TSC_C05320	1074	0.50	TTHA1310	1074	0.54	365	97%	0.41	227	53	62.19	14.52
TSC_C09520	1854	0.49	TTHA0586	1854	0.53	620	99%	0.41	271	53	43.71	8.55
TSC_C05310	930	0.47	TTHA1311	840	0.51	317	92%	0.41	183	63	57.73	19.87
TSC_C05650	729	0.46	TTHA0439	729	0.50	243	99%	0.41	105	25	43.21	10.29
TSC_C04100	627	0.50	TTHA1823	627	0.54	211	98%	0.41	131	37	62.09	17.54
TSC_C19110	741	0.53	TTHA0453	741	0.57	250	98%	0.41	134	44	53.60	17.60
TSC_C14450	1320	0.49	TTHA1078	1320	0.53	455	96%	0.41	328	109	72.09	23.96
TSC_C03410	468	0.47	TTHA0164	486	0.51	164	96%	0.41	74	21	45.12	12.80
TSC_C03850	1878	0.53	TTHA1720	2397	0.57	807	86%	0.41	346	94	42.87	11.65
TSC_C16970	705	0.44	TTHA0859	702	0.48	234	99%	0.41	58	5	24.79	2.14
TSC_C25050	741	0.54	TTHA1971	750	0.58	262	94%	0.41	156	53	59.54	20.23
TSC_C20880	378	0.48	TTHA1398	378	0.52	127	98%	0.40	68	18	53.54	14.17
TSC_C05880	183	0.33	TTHA0418	183	0.37	60	98%	0.40	19	1	31.67	1.67
TSC_C20230	678	0.50	TTHA1347	681	0.54	232	97%	0.40	148	44	63.79	18.97
TSC_C19100	807	0.47	TTHA0454	807	0.51	275	97%	0.40	134	40	48.73	14.55
TSC_C00690	1479	0.52	TTHA1952	1590	0.56	576	87%	0.40	278	86	48.26	14.93
TSC_C00630	450	0.52	TTHA1944	444	0.56	154	95%	0.40	81	22	52.60	14.29
TSC_C01990	1035	0.52	TTHA1898	978	0.56	357	93%	0.40	220	77	61.62	21.57
TSC_C00570	414	0.49	TTHA1943	411	0.53	149	90%	0.40	119	41	79.87	27.52
TSC_C10020	1068	0.49	TTHA0629	1068	0.53	361	98%	0.40	228	66	63.16	18.28
TSC_C23260	933	0.44	TTHA1637	933	0.48	310	100%	0.40	106	19	34.19	6.13
TSC_C01230	396	0.50	TTHA0344	396	0.54	135	96%	0.40	50	17	37.04	12.59
TSC_C09690	507	0.40	TTHA1134	507	0.44	168	99%	0.40	58	16	34.52	9.52
TSC_C04160	1731	0.53	TTHA1817	1725	0.56	584	98%	0.40	254	59	43.49	10.10
TSC_C07200	1080	0.54	TTHA0799	1080	0.58	362	99%	0.40	155	34	42.82	9.39

TSC_C05550	1647	0.45	TTHA1430	1647	0.49	551	99%	0.39	234	68	42.47	12.34
TSC_C21520	1314	0.50	TTHA1618	1311	0.54	441	99%	0.39	202	39	45.80	8.84
TSC_C01240	1242	0.57	TTHA0343	1245	0.61	426	97%	0.39	225	69	52.82	16.20
TSC_C09000	834	0.53	TTHA0692	831	0.57	286	96%	0.39	159	50	55.59	17.48
TSC_C06360	882	0.41	TTHA0119	888	0.45	295	99%	0.39	117	22	39.66	7.46
TSC_C18780	1863	0.49	TTHA0380	1863	0.53	624	99%	0.39	278	56	44.55	8.97
TSC_C14630	381	0.45	TTHA0427	381	0.49	126	99%	0.39	65	18	51.59	14.29
TSC_C04030	387	0.48	TTHA1829	408	0.52	137	95%	0.39	68	23	49.64	16.79
TSC_C24380	972	0.45	TTHA0326	1071	0.49	359	94%	0.39	153	40	42.62	11.14
TSC_C11750	741	0.53	TTHA0668	741	0.57	253	97%	0.39	131	37	51.78	14.62
TSC_C03750	390	0.43	TTHA1732	390	0.47	130	98%	0.39	94	26	72.31	20.00
TSC_C18530	1779	0.53	TTHA1534	1797	0.57	609	97%	0.39	376	114	61.74	18.72
TSC_C02170	891	0.48	TTHA0205	897	0.52	302	98%	0.39	174	49	57.62	16.23
TSC_C06570	1296	0.52	TTHA0140	1278	0.56	438	97%	0.39	256	62	58.45	14.16
TSC_C23440	402	0.47	TTHA1846	402	0.51	136	97%	0.39	125	44	91.91	32.35
TSC_C14940	2019	0.53	TTHA1167	2001	0.57	690	97%	0.39	435	127	63.04	18.41
TSC_C15170	2133	0.52	TTHA1183	2088	0.56	729	96%	0.39	405	102	55.56	13.99
TSC_C24780	1179	0.49	TTHA1923	1179	0.53	402	97%	0.39	282	82	70.15	20.40
TSC_C14920	2676	0.46	TTHA1169	2589	0.50	899	97%	0.38	368	75	40.93	8.34
TSC_C09630	1731	0.46	TTHA1140	1722	0.50	578	99%	0.38	176	41	30.45	7.09
TSC_C22170	942	0.47	TTHA1664	948	0.51	322	97%	0.38	137	30	42.55	9.32
TSC_C20890	1305	0.47	TTHA1399	1299	0.50	436	99%	0.38	194	40	44.50	9.17
TSC_C04150	1023	0.48	TTHA1818	1023	0.52	341	99%	0.38	128	26	37.54	7.62
TSC_C17770	975	0.55	TTHA0765	993	0.59	334	98%	0.38	186	49	55.69	14.67
TSC_C04120	780	0.50	TTHA1821	786	0.54	266	97%	0.38	190	50	71.43	18.80
TSC_C03390	714	0.54	TTHA0166	750	0.58	250	97%	0.38	143	30	57.20	12.00
TSC_C04020	684	0.46	TTHA1830	684	0.49	231	98%	0.38	140	43	60.61	18.61
TSC_C18990	366	0.37	TTHA1456	366	0.41	121	99%	0.38	52	13	42.98	10.74
TSC_C15520	567	0.48	TTHA0707	576	0.52	192	98%	0.38	44	10	22.92	5.21

TSC_C00470	636	0.52	TTHA0070	636	0.56	213	99%	0.38	126	27	59.15	12.68
TSC_C10200	888	0.52	TTHA1302	1050	0.56	355	89%	0.38	182	47	51.27	13.24
TSC_C14300	639	0.44	TTHA0578	636	0.48	217	97%	0.38	118	36	54.38	16.59
TSC_C15070	1338	0.49	TTHA1174	1341	0.53	455	98%	0.38	251	77	55.16	16.92
TSC_C03980	276	0.40	TTHA1832	285	0.44	95	96%	0.38	63	22	66.32	23.16
TSC_C12200	390	0.49	TTHA0981	324	0.53	131	88%	0.38	59	18	45.04	13.74
TSC_C21760	1311	0.49	TTHA1513	1311	0.52	441	99%	0.37	207	50	46.94	11.34
TSC_C22430	618	0.49	TTHA1691	633	0.53	210	98%	0.37	89	19	42.38	9.05
TSC_C01250	1131	0.50	TTHA0342	1131	0.54	381	98%	0.37	178	48	46.72	12.60
TSC_C23740	1356	0.52	TTHA1703	1356	0.56	453	99%	0.37	180	45	39.74	9.93
TSC_C05780	927	0.53	TTHA0408	927	0.56	318	96%	0.37	176	54	55.35	16.98
TSC_C22920	615	0.41	TTHA0994	594	0.45	217	91%	0.37	179	65	82.49	29.95
TSC_C08260	1650	0.44	TTHA0549	1647	0.48	551	99%	0.37	229	54	41.56	9.80
TSC_C08140	996	0.48	TTHA0545	996	0.52	334	99%	0.37	146	32	43.71	9.58
TSC_C23250	1116	0.49	TTHA1636	1131	0.53	382	97%	0.37	215	67	56.28	17.54
TSC_C07620	1038	0.50	TTHA1230	1038	0.53	345	100%	0.37	156	34	45.22	9.86
TSC_C00250	1848	0.47	TTHA0006	1848	0.51	618	99%	0.37	236	62	38.19	10.03
TSC_C10310	843	0.55	TTHA0950	825	0.58	287	96%	0.37	187	59	65.16	20.56
TSC_C17830	1293	0.51	TTHA1070	1236	0.55	445	94%	0.36	318	90	71.46	20.22
TSC_C25030	813	0.54	TTHA1969	810	0.57	271	99%	0.36	160	49	59.04	18.08
TSC_C13640	711	0.53	TTHA0903	603	0.57	246	87%	0.36	142	43	57.72	17.48
TSC_C09940	1224	0.43	TTHA0637	1224	0.47	409	99%	0.36	161	31	39.36	7.58
TSC_C17000	1104	0.50	TTHA0856	1104	0.54	370	99%	0.36	160	38	43.24	10.27
TSC_C09760	903	0.46	TTHA1126	903	0.50	303	99%	0.36	174	49	57.43	16.17
TSC_C03190	1275	0.52	TTHA1771	1272	0.56	428	99%	0.36	201	49	46.96	11.45
TSC_C10010	1251	0.49	TTHA0630	1251	0.53	418	99%	0.36	167	21	39.95	5.02
TSC_C17970	1125	0.46	TTHA1098	1188	0.50	405	94%	0.36	172	63	42.47	15.56
TSC_C15550	882	0.48	TTHA0704	894	0.52	297	99%	0.36	77	9	25.93	3.03
TSC_C01670	591	0.44	TTHA0328	591	0.48	196	99%	0.36	88	18	44.90	9.18

TSC_C04270	849	0.52	TTHA1809	861	0.56	291	97%	0.36	191	59	65.64	20.27
TSC_C15560	1194	0.51	TTHA0703	1116	0.55	407	94%	0.36	233	65	57.25	15.97
TSC_C02220	306	0.43	TTHA0245	306	0.47	101	99%	0.36	33	16	32.67	15.84
TSC_C13390	1593	0.48	TTHA0865	1548	0.52	534	98%	0.36	233	59	43.63	11.05
TSC_C09290	471	0.47	TTHA1042	471	0.50	158	98%	0.36	72	16	45.57	10.13
TSC_C22760	393	0.48	TTHA0401	393	0.52	131	98%	0.36	50	8	38.17	6.11
TSC_C21940	1428	0.49	TTHA1563	1428	0.52	477	99%	0.36	241	72	50.52	15.09
TSC_C18920	1302	0.49	TTHA1450	1314	0.52	448	97%	0.36	200	51	44.64	11.38
TSC_C20600	366	0.47	TTHA1375	375	0.50	124	98%	0.36	79	19	63.71	15.32
TSC_C00640	597	0.46	TTHA1946	597	0.50	200	98%	0.35	93	31	46.50	15.50
TSC_C20710	1152	0.49	TTHA1384	1152	0.52	383	100%	0.35	171	31	44.65	8.09
TSC_C17850	729	0.51	TTHA1295	729	0.54	243	99%	0.35	97	22	39.92	9.05
TSC_C21690	753	0.52	TTHA1505	747	0.56	260	95%	0.35	173	53	66.54	20.38
TSC_C23720	285	0.36	TTHA1701	285	0.39	96	97%	0.35	41	14	42.71	14.58
TSC_C17930	1218	0.48	TTHA1095	1257	0.52	419	98%	0.35	151	25	36.04	5.97
TSC_C00560	591	0.43	TTHA1942	585	0.46	197	98%	0.35	109	43	55.33	21.83
TSC_C07770	306	0.45	TTHA1649	288	0.48	116	81%	0.35	81	31	69.83	26.72
TSC_C01330	2010	0.47	TTHA0200	2013	0.50	675	99%	0.35	353	95	52.30	14.07
TSC_C13870	516	0.52	TTHA1032	489	0.56	173	96%	0.35	86	21	49.71	12.14
TSC_C17090	687	0.52	TTHA0847	687	0.55	228	100%	0.35	114	35	50.00	15.35
TSC_C14900	1053	0.52	TTHA0654	1053	0.56	352	99%	0.35	154	41	43.75	11.65
TSC_C24500	450	0.44	TTHA0043	453	0.47	154	96%	0.35	78	25	50.65	16.23
TSC_C08730	1959	0.50	TTHA1278	1959	0.53	655	99%	0.35	306	85	46.72	12.98
TSC_C01470	348	0.42	TTHA0348	348	0.46	119	96%	0.34	77	21	64.71	17.65
TSC_C11180	1218	0.49	TTHA1320	1215	0.52	435	92%	0.34	350	129	80.46	29.66
TSC_C12990	1077	0.48	TTHA0594	1077	0.52	362	99%	0.34	175	45	48.34	12.43
TSC_C04320	240	0.45	TTHA1803	243	0.48	82	96%	0.34	44	13	53.66	15.85
TSC_C16980	558	0.39	TTHA0858	558	0.42	186	99%	0.34	60	16	32.26	8.60
TSC_C20260	1884	0.51	TTHA1492	1875	0.54	631	99%	0.34	251	51	39.78	8.08

TSC_C13330	1269	0.49	TTHA0875	1266	0.53	427	98%	0.34	212	61	49.65	14.29
TSC_C21600	2397	0.49	TTHA1500	2394	0.53	804	99%	0.34	331	89	41.17	11.07
TSC_C07740	1191	0.51	TTHA0559	1191	0.55	397	99%	0.34	202	43	50.88	10.83
TSC_C18360	2388	0.47	TTHA0770	2388	0.50	796	100%	0.34	297	52	37.31	6.53
TSC_C20800	1299	0.46	TTHA1391	1296	0.49	434	99%	0.34	186	45	42.86	10.37
TSC_C18190	426	0.53	TTHA1254	426	0.57	144	97%	0.34	71	19	49.31	13.19
TSC_C09270	876	0.49	TTHA1040	876	0.52	299	97%	0.34	196	59	65.55	19.73
TSC_C02450	804	0.51	TTHA0223	792	0.55	275	96%	0.34	185	64	67.27	23.27
TSC_C01160	444	0.37	TTHA0255	444	0.41	148	99%	0.34	73	15	49.32	10.14
TSC_C01220	906	0.49	TTHA0345	906	0.53	305	98%	0.34	138	34	45.25	11.15
TSC_C07580	1671	0.51	TTHA1234	1668	0.54	563	98%	0.34	239	54	42.45	9.59
TSC_C06130	705	0.53	TTHA0422	705	0.57	239	97%	0.34	131	36	54.81	15.06
TSC_C01690	1296	0.50	TTHA0252	1296	0.54	433	99%	0.34	243	75	56.12	17.32
TSC_C09950	399	0.44	TTHA0636	360	0.47	136	91%	0.34	83	28	61.03	20.59
TSC_C20420	633	0.44	TTHA1474	633	0.48	223	93%	0.33	128	43	57.40	19.28
TSC_C22220	768	0.49	TTHA1670	768	0.53	255	100%	0.33	94	22	36.86	8.63
TSC_C06370	906	0.48	TTHA0120	906	0.51	301	100%	0.33	108	23	35.88	7.64
TSC_C04340	1377	0.49	TTHA0040	1350	0.52	459	99%	0.33	197	45	42.92	9.80
TSC_C18710	1314	0.49	TTHA1442	1332	0.53	448	98%	0.33	224	54	50.00	12.05
TSC_C09870	675	0.46	TTHA1115	675	0.49	228	98%	0.33	127	40	55.70	17.54
TSC_C22740	1908	0.47	TTHA1586	1905	0.50	638	99%	0.33	214	39	33.54	6.11
TSC_C04310	360	0.47	TTHA1804	378	0.51	126	96%	0.33	43	6	34.13	4.76
TSC_C05160	1704	0.54	TTHA1323	1638	0.57	585	95%	0.33	369	109	63.08	18.63
TSC_C23650	483	0.41	TTHA1932	483	0.44	160	99%	0.33	79	22	49.38	13.75
TSC_C05170	858	0.46	TTHA1321	858	0.49	285	100%	0.33	98	26	34.39	9.12
TSC_C01440	1152	0.50	TTHA0351	1137	0.54	389	98%	0.33	217	52	55.78	13.37
TSC_C22310	186	0.43	TTHA1679	186	0.46	61	98%	0.33	20	2	32.79	3.28
TSC_C09770	555	0.45	TTHA1125	555	0.48	185	99%	0.33	69	20	37.30	10.81
TSC_C09410	288	0.48	TTHA0907	288	0.51	96	98%	0.33	70	23	72.92	23.96

TSC_C23490	924	0.46	TTHA1856	924	0.50	309	99%	0.33	130	30	42.07	9.71
TSC_C08830	3300	0.50	TTHA0729	3300	0.53	1116	98%	0.33	546	146	48.92	13.08
TSC_C22050	1275	0.49	TTHA1576	1275	0.52	425	100%	0.33	175	34	41.18	8.00
TSC_C19010	435	0.41	TTHA1457	444	0.44	148	98%	0.32	51	14	34.46	9.46
TSC_C23680	1620	0.48	TTHA0162	1611	0.51	546	98%	0.32	244	69	44.69	12.64
TSC_C09670	480	0.42	TTHA1136	471	0.45	167	93%	0.32	112	37	67.07	22.16
TSC_C02830	228	0.43	TTHA1784	354	0.47	117	77%	0.32	39	10	33.33	8.55
TSC_C18430	2856	0.53	TTHA0777	2865	0.56	1020	93%	0.32	782	269	76.67	26.37
TSC_C01830	1119	0.51	TTHA1916	1119	0.54	372	100%	0.32	150	33	40.32	8.87
TSC_C24360	1143	0.52	TTHA0324	1146	0.55	384	99%	0.32	175	50	45.57	13.02
TSC_C18900	996	0.36	TTHA1364	1014	0.39	353	94%	0.32	321	103	90.93	29.18
TSC_C23050	1059	0.55	TTHA1447	1038	0.58	354	98%	0.32	229	71	64.69	20.06
TSC_C14030	1524	0.49	TTHA0617	1521	0.52	513	99%	0.32	213	51	41.52	9.94
TSC_C13820	858	0.43	TTHA1028	858	0.46	286	99%	0.31	106	25	37.06	8.74
TSC_C06400	897	0.50	TTHA0123	954	0.54	325	94%	0.31	219	64	67.38	19.69
TSC_C07660	504	0.47	TTHA1223	762	0.50	256	78%	0.31	127	42	49.61	16.41
TSC_C24490	297	0.38	TTHA0042	297	0.41	99	98%	0.31	30	5	30.30	5.05
TSC_C02420	429	0.50	TTHA0226	426	0.53	145	97%	0.31	61	13	42.07	8.97
TSC_C16960	591	0.43	TTHA0860	591	0.46	196	99%	0.31	55	10	28.06	5.10
TSC_C19140	972	0.48	TTHA0406	975	0.52	326	99%	0.31	126	23	38.65	7.06
TSC_C23580	960	0.54	TTHA0315	966	0.57	329	97%	0.31	202	71	61.40	21.58
TSC_C22910	936	0.50	TTHA1619	936	0.53	317	98%	0.31	182	54	57.41	17.03
TSC_C24930	456	0.53	TTHA1961	465	0.56	168	89%	0.31	108	40	64.29	23.81
TSC_C17180	1437	0.50	TTHA0837	1437	0.53	479	100%	0.31	183	38	38.20	7.93
TSC_C08800	672	0.50	TTHA1271	672	0.53	223	100%	0.31	84	18	37.67	8.07
TSC_C10190	1506	0.54	TTHA1304	1470	0.57	516	96%	0.31	311	89	60.27	17.25
TSC_C22490	411	0.44	TTHA1697	408	0.47	138	97%	0.31	34	6	24.64	4.35
TSC_C18400	693	0.41	TTHA0774	687	0.44	248	91%	0.31	117	38	47.18	15.32
TSC_C14340	744	0.52	TTHA0573	744	0.55	253	97%	0.31	152	35	60.08	13.83

TSC_C03820	1122	0.50	TTHA1723	1122	0.53	382	97%	0.31	214	75	56.02	19.63
TSC_C14810	1236	0.47	TTHA0688	1260	0.50	421	98%	0.31	138	28	32.78	6.65
TSC_C05410	1287	0.49	TTHA0503	1257	0.52	430	98%	0.31	201	60	46.74	13.95
TSC_C24520	1158	0.49	TTHA0046	1158	0.52	386	99%	0.30	147	32	38.08	8.29
TSC_C23510	789	0.53	TTHA1859	768	0.56	265	97%	0.30	127	45	47.92	16.98
TSC_C14100	3108	0.49	TTHA0612	3087	0.52	1049	98%	0.30	464	111	44.23	10.58
TSC_C07070	984	0.47	TTHA0536	984	0.50	327	100%	0.30	134	33	40.98	10.09
TSC_C13020	1230	0.44	TTHA1144	1233	0.47	420	97%	0.30	154	28	36.67	6.67
TSC_C00820	306	0.40	TTHA0272	306	0.43	101	99%	0.30	21	4	20.79	3.96
TSC_C06460	546	0.40	TTHA0130	546	0.43	181	99%	0.30	71	15	39.23	8.29
TSC_C07800	636	0.50	TTHA0563	636	0.53	213	99%	0.30	90	28	42.25	13.15
TSC_C06080	645	0.46	TTHA0081	639	0.49	214	99%	0.30	102	20	47.66	9.35
TSC_C06410	954	0.47	TTHA0124	927	0.50	319	98%	0.30	94	18	29.47	5.64
TSC_C05920	1281	0.48	TTHA0097	1284	0.51	438	97%	0.30	231	63	52.74	14.38
TSC_C03300	1050	0.55	TTHA0174	1053	0.58	354	98%	0.30	196	60	55.37	16.95
TSC_C14090	285	0.48	TTHA0613	285	0.51	95	98%	0.30	61	14	64.21	14.74
TSC_C23590	1980	0.47	TTHA1875	1980	0.50	662	99%	0.30	272	71	41.09	10.73
TSC_C18030	927	0.47	TTHA1104	843	0.50	312	94%	0.30	128	30	41.03	9.62
TSC_C05590	726	0.56	TTHA1426	723	0.59	270	87%	0.30	181	60	67.04	22.22
TSC_C03990	423	0.51	TTHA1571	399	0.54	159	82%	0.30	127	44	79.87	27.67
TSC_C10420	891	0.48	TTHA0957	900	0.50	301	98%	0.30	125	28	41.53	9.30
TSC_C23080	843	0.46	TTHA0383	840	0.49	284	98%	0.30	137	40	48.24	14.08
TSC_C23520	810	0.51	TTHA1860	807	0.54	276	97%	0.30	183	52	66.30	18.84
TSC_C22390	342	0.44	TTHA1687	342	0.47	113	99%	0.30	30	5	26.55	4.42
TSC_C22400	282	0.39	TTHA1688	282	0.42	93	99%	0.29	19	0	20.43	0.00
TSC_C01350	618	0.49	TTHA0198	612	0.51	210	97%	0.29	124	32	59.05	15.24
TSC_C21080	579	0.40	TTHA1418	573	0.43	194	98%	0.29	101	27	52.06	13.92
TSC_C22620	1227	0.49	TTHA0222	1227	0.52	411	99%	0.29	175	38	42.58	9.25
TSC_C20160	420	0.48	TTHA0460	411	0.51	140	97%	0.29	81	27	57.86	19.29

TSC_C09150	1716	0.47	TTHA0699	1716	0.50	575	99%	0.29	207	48	36.00	8.35
TSC_C13080	1728	0.48	TTHA1150	1728	0.51	584	98%	0.29	334	76	57.19	13.01
TSC_C03710	417	0.48	TTHA1736	417	0.51	139	99%	0.29	43	7	30.94	5.04
TSC_C22630	990	0.56	TTHA1842	990	0.59	334	98%	0.29	171	48	51.20	14.37
TSC_C04420	930	0.50	TTHA1205	996	0.53	376	82%	0.29	345	132	91.76	35.11
TSC_C23390	1320	0.41	TTHA1650	1320	0.44	440	100%	0.29	171	40	38.86	9.09
TSC_C15010	1974	0.50	TTHA1158	1974	0.52	663	99%	0.29	344	75	51.89	11.31
TSC_C24140	804	0.46	TTHA1886	786	0.49	270	97%	0.29	121	28	44.81	10.37
TSC_C03280	543	0.46	TTHA0176	474	0.49	186	89%	0.29	99	37	53.23	19.89
TSC_C01290	1200	0.48	TTHA1798	1191	0.51	410	97%	0.29	205	58	50.00	14.15
TSC_C05890	522	0.48	TTHA0419	537	0.51	180	97%	0.29	73	21	40.56	11.67
TSC_C24710	1272	0.50	TTHA1928	1242	0.53	425	98%	0.29	211	50	49.65	11.76
TSC_C19640	966	0.51	TTHA0477	963	0.54	321	100%	0.29	158	41	49.22	12.77
TSC_C14560	1056	0.48	TTHA1089	1059	0.51	353	99%	0.29	134	23	37.96	6.52
TSC_C00330	915	0.49	TTHA0347	915	0.52	305	99%	0.29	112	11	36.72	3.61
TSC_C18200	210	0.46	TTHA1255	210	0.48	70	97%	0.29	33	7	47.14	10.00
TSC_C19680	984	0.46	TTHA0473	984	0.49	329	99%	0.29	122	19	37.08	5.78
TSC_C18980	396	0.35	TTHA1455	396	0.38	131	99%	0.29	54	16	41.22	12.21
TSC_C20540	756	0.51	TTHA1369	756	0.54	261	96%	0.28	187	59	71.65	22.61
TSC_C21060	543	0.44	TTHA1416	561	0.47	189	96%	0.28	96	29	50.79	15.34
TSC_C09310	399	0.45	TTHA1044	399	0.47	135	97%	0.28	83	23	61.48	17.04
TSC_C07360	2502	0.52	TTHA0815	2502	0.55	891	93%	0.28	536	182	60.16	20.43
TSC_C17960	2025	0.49	TTHA1097	2031	0.52	678	99%	0.28	312	60	46.02	8.85
TSC_C00080	798	0.42	TTHA1850	744	0.45	304	81%	0.28	253	101	83.22	33.22
TSC_C13910	1350	0.49	TTHA1060	1353	0.51	460	97%	0.28	230	68	50.00	14.78
TSC_C07520	1080	0.47	TTHA1241	1080	0.50	370	97%	0.28	206	56	55.68	15.14
TSC_C17240	954	0.47	TTHA0600	954	0.50	317	100%	0.28	142	41	44.79	12.93
TSC_C03480	1005	0.43	TTHA1760	999	0.46	338	98%	0.28	186	63	55.03	18.64
TSC_C02040	1080	0.50	TTHA0269	1068	0.53	366	97%	0.28	230	66	62.84	18.03

TSC_C15390	345	0.49	TTHA0713	318	0.51	126	84%	0.28	75	24	59.52	19.05
TSC_C17780	1062	0.44	TTHA0764	1062	0.47	353	100%	0.28	155	36	43.91	10.20
TSC_C00750	915	0.45	TTHA1956	915	0.48	305	99%	0.28	103	21	33.77	6.89
TSC_C04570	1035	0.51	TTHA0016	1029	0.54	351	97%	0.28	219	63	62.39	17.95
TSC_C11810	195	0.45	TTHA0675	195	0.47	65	97%	0.28	44	17	67.69	26.15
TSC_C22070	945	0.50	TTHA1655	939	0.52	320	97%	0.28	175	35	54.69	10.94
TSC_C09680	1689	0.43	TTHA1135	1689	0.46	562	100%	0.28	232	46	41.28	8.19
TSC_C22280	339	0.42	TTHA1676	339	0.45	114	97%	0.27	38	3	33.33	2.63
TSC_C07780	2466	0.50	TTHA0561	2469	0.53	840	98%	0.27	450	162	53.57	19.29
TSC_C01960	867	0.51	TTHA1901	879	0.54	294	98%	0.27	134	37	45.58	12.59
TSC_C03110	1098	0.54	TTHA1778	1098	0.56	379	96%	0.27	227	85	59.89	22.43
TSC_C23140	429	0.50	TTHA1625	429	0.53	144	98%	0.27	51	15	35.42	10.42
TSC_C00320	639	0.47	TTHA1012	645	0.50	219	97%	0.27	151	50	68.95	22.83
TSC_C23820	1053	0.47	TTHA1710	1053	0.50	351	99%	0.27	162	45	46.15	12.82
TSC_C18580	1725	0.46	TTHA1532	1713	0.49	578	99%	0.27	270	69	46.71	11.94
TSC_C24720	1122	0.52	TTHA1927	1113	0.55	377	98%	0.27	232	57	61.54	15.12
TSC_C10220	1122	0.47	TTHA1301	1140	0.50	380	99%	0.27	158	39	41.58	10.26
TSC_C23040	1323	0.49	TTHA1448	1335	0.52	445	99%	0.27	164	30	36.85	6.74
TSC_C08990	732	0.52	TTHA0691	729	0.55	247	98%	0.27	129	27	52.23	10.93
TSC_C12650	1434	0.53	TTHA0996	1428	0.56	480	99%	0.27	185	42	38.54	8.75
TSC_C24350	714	0.50	TTHA0323	726	0.53	242	98%	0.27	141	36	58.26	14.88
TSC_C23150	591	0.51	TTHA1626	591	0.54	196	99%	0.27	26	3	13.27	1.53
TSC_C19700	1359	0.44	TTHA0471	1347	0.47	465	96%	0.27	289	74	62.15	15.91
TSC_C21300	1206	0.48	TTHA1597	1194	0.51	405	98%	0.27	221	51	54.57	12.59
TSC_C23730	399	0.49	TTHA1702	399	0.52	132	99%	0.27	61	18	46.21	13.64
TSC_C06260	531	0.45	TTHA0517	537	0.48	194	90%	0.27	128	47	65.98	24.23
TSC_C05640	1419	0.48	TTHA0438	1407	0.51	474	99%	0.27	190	47	40.08	9.92
TSC_C05140	1245	0.43	TTHA1325	1224	0.46	414	99%	0.27	191	47	46.14	11.35
TSC_C01140	885	0.47	TTHA0257	891	0.50	299	98%	0.27	122	32	40.80	10.70

TSC_C00150	1209	0.49	TTHA0193	1209	0.52	409	98%	0.27	213	46	52.08	11.25
TSC_C23480	1839	0.49	TTHA1855	1815	0.52	618	98%	0.27	287	66	46.44	10.68
TSC_C01570	486	0.39	TTHA0337	405	0.42	162	90%	0.27	74	20	45.68	12.35
TSC_C10410	1377	0.54	TTHA0956	1497	0.57	508	94%	0.26	309	91	60.83	17.91
TSC_C03970	1143	0.53	TTHA1833	1143	0.55	390	97%	0.26	229	68	58.72	17.44
TSC_C07140	1221	0.45	TTHA0794	1221	0.48	406	100%	0.26	160	33	39.41	8.13
TSC_C04190	1032	0.47	TTHA1816	1038	0.50	345	99%	0.26	132	23	38.26	6.67
TSC_C24630	525	0.51	TTHA0059	672	0.54	223	87%	0.26	62	15	27.80	6.73
TSC_C19090	465	0.53	TTHA0455	450	0.55	157	96%	0.26	70	18	44.59	11.46
TSC_C23160	777	0.46	TTHA1627	777	0.49	264	97%	0.26	126	43	47.73	16.29
TSC_C08270	762	0.50	TTHA0550	762	0.53	255	99%	0.26	133	31	52.16	12.16
TSC_C08410	1014	0.48	TTHA1211	1014	0.51	338	99%	0.26	123	29	36.39	8.58
TSC_C19310	924	0.49	TTHA1549	924	0.52	308	99%	0.26	150	34	48.70	11.04
TSC_C05530	876	0.46	TTHA1433	873	0.49	292	99%	0.26	108	26	36.99	8.90
TSC_C03200	969	0.54	TTHA1770	900	0.57	325	95%	0.26	182	57	56.00	17.54
TSC_C22870	1065	0.47	TTHA0388	1068	0.50	355	100%	0.26	136	27	38.31	7.61
TSC_C08780	1737	0.49	TTHA1273	1737	0.52	578	100%	0.25	189	30	32.70	5.19
TSC_C02310	870	0.52	TTHA0237	870	0.54	290	99%	0.25	117	28	40.34	9.66
TSC_C03950	1332	0.55	TTHA1835	1332	0.57	445	99%	0.25	182	56	40.90	12.58
TSC_C06480	480	0.47	TTHA0132	480	0.50	160	99%	0.25	54	12	33.75	7.50
TSC_C07030	1485	0.46	TTHA0532	1272	0.48	499	91%	0.25	163	40	32.67	8.02
TSC_C03320	1062	0.57	TTHA0173	1062	0.59	354	99%	0.25	150	30	42.37	8.47
TSC_C17110	549	0.48	TTHA0844	549	0.51	197	91%	0.25	133	40	67.51	20.30
TSC_C05270	1197	0.48	TTHA1315	1200	0.51	409	97%	0.25	230	54	56.23	13.20
TSC_C18100	954	0.45	TTHA0824	945	0.47	320	98%	0.25	127	34	39.69	10.63
TSC_C22240	1317	0.45	TTHA1672	1317	0.48	438	100%	0.25	194	32	44.29	7.31
TSC_C03550	1194	0.48	TTHA1755	1188	0.51	397	99%	0.25	188	52	47.36	13.10
TSC_C00070	1431	0.48	TTHA1785	1431	0.50	478	99%	0.25	188	37	39.33	7.74
TSC_C05730	1935	0.47	TTHA0448	1935	0.50	650	99%	0.25	251	48	38.62	7.38

TSC_C08880	1293	0.46	TTHA0734	1338	0.49	451	97%	0.25	215	63	47.67	13.97
TSC_C01550	2667	0.48	TTHA0364	2670	0.51	894	99%	0.25	381	102	42.62	11.41
TSC_C22180	630	0.45	TTHA1665	630	0.47	209	100%	0.25	70	14	33.49	6.70
TSC_C03460	519	0.44	TTHA1762	519	0.47	178	96%	0.24	84	31	47.19	17.42
TSC_C07260	681	0.51	TTHA0805	666	0.53	230	97%	0.24	126	32	54.78	13.91
TSC_C21000	549	0.43	TTHA1410	558	0.46	186	98%	0.24	105	29	56.45	15.59
TSC_C00180	480	0.47	TTHA1789	474	0.49	161	97%	0.24	74	19	45.96	11.80
TSC_C13190	1194	0.50	TTHA0891	1194	0.53	398	99%	0.24	161	27	40.45	6.78
TSC_C00790	4941	0.50	TTHA0180	6204	0.53	2093	87%	0.24	776	198	37.08	9.46
TSC_C15700	606	0.43	TTHA0832	606	0.46	203	99%	0.24	138	35	67.98	17.24
TSC_C20080	1737	0.52	TTHA0114	1758	0.55	590	98%	0.24	331	105	56.10	17.80
TSC_C01460	432	0.47	TTHA0349	432	0.50	145	98%	0.24	81	27	55.86	18.62
TSC_C17200	762	0.49	TTHA0835	753	0.51	256	98%	0.24	126	26	49.22	10.16
TSC_C03800	759	0.49	TTHA1726	753	0.52	275	90%	0.24	209	83	76.00	30.18
TSC_C07410	1743	0.46	TTHA0819	1674	0.48	602	94%	0.24	348	136	57.81	22.59
TSC_C05700	195	0.42	TTHA0445	234	0.44	77	90%	0.24	32	10	41.56	12.99
TSC_C24030	381	0.51	TTHA1880	381	0.53	127	98%	0.24	60	18	47.24	14.17
TSC_C08050	867	0.48	TTHA0537	867	0.50	288	100%	0.24	93	17	32.29	5.90
TSC_C06850	1731	0.49	TTHA0520	1731	0.51	577	100%	0.24	242	57	41.94	9.88
TSC_C04460	195	0.42	TTHA0026	210	0.44	71	92%	0.24	30	12	42.25	16.90
TSC_C10330	1566	0.53	TTHA0952	1566	0.56	522	100%	0.24	223	44	42.72	8.43
TSC_C19530	1332	0.50	TTHA0484	1431	0.52	484	94%	0.24	300	68	61.98	14.05
TSC_C06390	909	0.41	TTHA0122	909	0.44	303	99%	0.24	83	13	27.39	4.29
TSC_C03140	831	0.53	TTHA1775	831	0.55	282	97%	0.24	133	28	47.16	9.93
TSC_C21580	2001	0.52	TTHA1498	1977	0.54	671	98%	0.23	252	61	37.56	9.09
TSC_C08400	513	0.45	TTHA1212	513	0.47	170	99%	0.23	79	13	46.47	7.65
TSC_C01980	414	0.44	TTHA1899	423	0.46	142	97%	0.23	83	26	58.45	18.31
TSC_C05440	1563	0.44	TTHA0506	1563	0.47	522	99%	0.23	169	31	32.38	5.94
TSC_C21310	981	0.48	TTHA1598	999	0.51	332	99%	0.23	160	35	48.19	10.54

TSC_C02460	381	0.46	TTHA1713	390	0.48	131	96%	0.23	72	19	54.96	14.50
TSC_C01370	1080	0.48	TTHA0196	1080	0.50	361	99%	0.23	160	37	44.32	10.25
TSC_C01520	1002	0.52	TTHA0367	1002	0.55	337	99%	0.23	190	52	56.38	15.43
TSC_C07450	1947	0.48	TTHA1248	1947	0.50	650	100%	0.23	222	50	34.15	7.69
TSC_C20270	1860	0.48	TTHA1491	1848	0.50	619	100%	0.23	185	42	29.89	6.79
TSC_C18630	405	0.49	TTHA1546	414	0.52	139	97%	0.23	57	14	41.01	10.07
TSC_C03040	1887	0.47	TTHA0641	1890	0.49	643	98%	0.23	510	173	79.32	26.91
TSC_C19360	288	0.42	TTHA1554	288	0.44	96	98%	0.23	30	7	31.25	7.29
TSC_C14140	372	0.42	TTHA0606	363	0.44	123	98%	0.23	50	11	40.65	8.94
TSC_C10240	552	0.48	TTHA0863	549	0.51	186	98%	0.22	123	33	66.13	17.74
TSC_C18470	1323	0.48	TTHA1308	1398	0.50	466	97%	0.22	158	22	33.91	4.72
TSC_C06470	903	0.50	TTHA0131	885	0.52	309	96%	0.22	205	73	66.34	23.62
TSC_C09350	984	0.48	TTHA1048	978	0.50	331	98%	0.22	107	12	32.33	3.63
TSC_C00720	282	0.37	TTHA1953	282	0.39	93	99%	0.22	32	6	34.41	6.45
TSC_C06900	1317	0.52	TTHA0525	1317	0.54	441	99%	0.22	237	62	53.74	14.06
TSC_C07730	471	0.44	TTHA1216	471	0.46	157	99%	0.22	115	37	73.25	23.57
TSC_C03540	207	0.41	TTHA1756	204	0.44	71	93%	0.22	52	12	73.24	16.90
TSC_C13620	996	0.44	TTHA0905	996	0.47	334	99%	0.22	127	25	38.02	7.49
TSC_C24670	1173	0.52	TTHA1938	1164	0.54	390	99%	0.22	160	41	41.03	10.51
TSC_C05990	2349	0.51	TTHA0090	2352	0.53	793	99%	0.22	344	99	43.38	12.48
TSC_C04940	1230	0.52	TTHA1353	1221	0.54	416	98%	0.22	231	72	55.53	17.31
TSC_C02510	753	0.45	TTHA1714	750	0.47	256	97%	0.22	141	52	55.08	20.31
TSC_C20950	429	0.47	TTHA1405	429	0.50	143	99%	0.21	61	19	42.66	13.29
TSC_C02940	1113	0.43	TTHA0648	1116	0.45	447	79%	0.21	382	128	85.46	28.64
TSC_C21050	804	0.45	TTHA1415	807	0.47	272	98%	0.21	118	32	43.38	11.76
TSC_C04470	2313	0.50	TTHA0024	2322	0.52	786	98%	0.21	424	136	53.94	17.30
TSC_C19040	957	0.43	TTHA1460	957	0.45	318	100%	0.21	119	33	37.42	10.38
TSC_C21720	366	0.47	TTHA1508	366	0.49	124	97%	0.21	49	7	39.52	5.65
TSC_C09740	897	0.49	TTHA1128	885	0.51	305	97%	0.21	186	54	60.98	17.70

TSC_C24010	609	0.49	TTHA1878	663	0.51	221	95%	0.21	80	19	36.20	8.60
TSC_C06670	180	0.38	TTHA1026	180	0.40	59	98%	0.21	6	3	10.17	5.08
TSC_C12210	477	0.42	TTHA0982	405	0.45	159	90%	0.21	58	16	36.48	10.06
TSC_C03900	1293	0.47	TTHA1840	1296	0.50	432	99%	0.21	173	38	40.05	8.80
TSC_C08060	1137	0.49	TTHA0538	1176	0.51	391	98%	0.21	96	17	24.55	4.35
TSC_C13180	1734	0.50	TTHA0892	1734	0.52	577	100%	0.21	223	34	38.65	5.89
TSC_C20340	426	0.47	TTHA1484	414	0.49	141	98%	0.21	62	19	43.97	13.48
TSC_C23550	201	0.29	TTHA1863	201	0.31	66	99%	0.20	39	12	59.09	18.18
TSC_C02250	444	0.45	TTHA0242	447	0.47	148	99%	0.20	33	8	22.30	5.41
TSC_C00810	1629	0.48	TTHA0271	1632	0.50	543	100%	0.20	125	25	23.02	4.60
TSC_C13310	897	0.51	TTHA0877	744	0.53	300	90%	0.20	118	27	39.33	9.00
TSC_C08390	1689	0.49	TTHA1213	1689	0.51	562	100%	0.20	204	41	36.30	7.30
TSC_C21930	663	0.48	TTHA1562	672	0.50	225	98%	0.20	103	22	45.78	9.78
TSC_C08680	675	0.47	TTHA1283	675	0.49	224	100%	0.20	104	27	46.43	12.05
TSC_C17030	573	0.52	TTHA0853	582	0.54	209	90%	0.20	170	66	81.34	31.58
TSC_C09590	567	0.42	TTHA0593	567	0.44	188	99%	0.20	77	20	40.96	10.64
TSC_C20650	1005	0.46	TTHA1378	1005	0.48	334	100%	0.20	123	22	36.83	6.59
TSC_C04050	837	0.44	TTHA1827	846	0.46	283	98%	0.19	136	32	48.06	11.31
TSC_C17840	561	0.44	TTHA1294	561	0.46	186	99%	0.19	81	20	43.55	10.75
TSC_C22420	291	0.44	TTHA1690	291	0.46	96	99%	0.19	29	7	30.21	7.29
TSC_C13380	369	0.45	TTHA0868	369	0.47	122	99%	0.19	50	5	40.98	4.10
TSC_C23090	1068	0.51	TTHA1620	1065	0.53	361	98%	0.19	185	52	51.25	14.40
TSC_C01510	630	0.46	TTHA0368	633	0.48	212	98%	0.19	156	47	73.58	22.17
TSC_C21980	627	0.56	TTHA1567	624	0.58	222	93%	0.19	122	43	54.95	19.37
TSC_C17130	1044	0.51	TTHA0842	1056	0.53	354	98%	0.19	128	36	36.16	10.17
TSC_C13970	231	0.40	TTHA0624	246	0.42	82	94%	0.19	61	15	74.39	18.29
TSC_C04980	1755	0.49	TTHA1342	1737	0.51	598	97%	0.19	296	80	49.50	13.38
TSC_C22160	357	0.43	TTHA1663	357	0.45	118	99%	0.19	36	6	30.51	5.08
TSC_C09880	939	0.48	TTHA1114	939	0.50	318	98%	0.19	185	45	58.18	14.15

TSC_C21240	579	0.54	TTHA1591	588	0.56	211	90%	0.19	127	45	60.19	21.33
TSC_C04880	609	0.51	TTHA1359	609	0.53	209	96%	0.19	100	30	47.85	14.35
TSC_C01600	384	0.50	TTHA0335	378	0.52	130	96%	0.18	74	24	56.92	18.46
TSC_C04240	4593	0.50	TTHA1812	4575	0.51	1536	99%	0.18	477	107	31.05	6.97
TSC_C13500	1119	0.51	TTHA0940	1146	0.53	385	97%	0.18	187	45	48.57	11.69
TSC_C04010	2649	0.52	TTHA1831	2649	0.54	891	99%	0.18	376	89	42.20	9.99
TSC_C22460	1221	0.40	TTHA0251	1221	0.42	406	100%	0.18	79	9	19.46	2.22
TSC_C14370	408	0.46	TTHA0571	411	0.48	139	97%	0.18	64	14	46.04	10.07
TSC_C15220	1161	0.49	TTHA1187	1131	0.51	390	97%	0.18	263	90	67.44	23.08
TSC_C24330	576	0.48	TTHA0321	579	0.50	192	99%	0.18	92	21	47.92	10.94
TSC_C21180	1650	0.50	TTHA1466	1653	0.52	555	99%	0.18	209	39	37.66	7.03
TSC_C24070	807	0.47	TTHA1159	735	0.49	302	82%	0.18	212	89	70.20	29.47
TSC_C21410	486	0.48	TTHA1608	456	0.50	167	92%	0.18	74	18	44.31	10.78
TSC_C04990	1827	0.48	TTHA1341	1803	0.50	613	98%	0.18	280	83	45.68	13.54
TSC_C21700	1179	0.54	TTHA1506	1173	0.56	404	96%	0.18	192	65	47.52	16.09
TSC_C06910	1425	0.49	TTHA0526	1425	0.51	476	99%	0.17	175	46	36.76	9.66
TSC_C17870	1884	0.47	TTHA1298	1857	0.49	634	98%	0.17	306	66	48.26	10.41
TSC_C11860	759	0.53	TTHA0211	762	0.54	255	99%	0.17	126	25	49.41	9.80
TSC_C00670	747	0.54	TTHA1950	750	0.56	250	99%	0.17	144	41	57.60	16.40
TSC_C03260	540	0.56	TTHA0178	540	0.57	187	95%	0.17	109	39	58.29	20.86
TSC_C07630	606	0.47	TTHA1229	606	0.49	201	100%	0.17	78	22	38.81	10.95
TSC_C20360	414	0.42	TTHA1480	414	0.43	138	99%	0.17	68	16	49.28	11.59
TSC_C21480	522	0.47	TTHA1614	537	0.49	178	98%	0.17	77	11	43.26	6.18
TSC_C04810	1485	0.50	TTHA0432	1482	0.52	496	99%	0.17	187	36	37.70	7.26
TSC_C06270	696	0.46	TTHA0518	693	0.47	232	99%	0.17	118	34	50.86	14.66
TSC_C22320	549	0.43	TTHA1680	549	0.44	182	99%	0.17	49	9	26.92	4.95
TSC_C07210	1035	0.50	TTHA0800	1035	0.52	345	99%	0.17	137	28	39.71	8.12
TSC_C08290	198	0.39	TTHA0552	198	0.41	65	98%	0.17	19	7	29.23	10.77
TSC_C18210	1200	0.48	TTHA0789	1158	0.50	400	98%	0.17	139	26	34.75	6.50

TSC_C06530	945	0.54	TTHA0136	948	0.56	319	98%	0.16	154	35	48.28	10.97
TSC_C06180	585	0.51	TTHA0429	585	0.53	195	99%	0.16	83	16	42.56	8.21
TSC_C20870	294	0.42	TTHA1397	321	0.44	106	95%	0.16	42	10	39.62	9.43
TSC_C02030	921	0.53	TTHA0267	921	0.54	309	99%	0.16	179	48	57.93	15.53
TSC_C02150	1143	0.52	TTHA0203	1140	0.54	390	97%	0.16	173	47	44.36	12.05
TSC_C04130	1185	0.55	TTHA1820	1185	0.56	397	99%	0.16	178	53	44.84	13.35
TSC_C18070	492	0.43	TTHA0827	492	0.44	163	99%	0.16	67	22	41.10	13.50
TSC_C16950	789	0.44	TTHA0861	771	0.46	265	97%	0.16	83	19	31.32	7.17
TSC_C09850	1986	0.47	TTHA1117	1986	0.49	667	99%	0.16	282	66	42.28	9.90
TSC_C19650	222	0.45	TTHA0476	198	0.46	77	87%	0.16	34	8	44.16	10.39
TSC_C00120	357	0.49	TTHA0190	357	0.51	118	99%	0.16	45	10	38.14	8.47
TSC_C07430	1878	0.50	TTHA1250	1884	0.52	629	99%	0.16	236	54	37.52	8.59
TSC_C23300	1077	0.44	TTHA1641	1083	0.45	377	95%	0.16	242	90	64.19	23.87
TSC_C24050	1764	0.50	TTHA0741	1833	0.51	735	77%	0.16	499	233	67.89	31.70
TSC_C03910	1407	0.46	TTHA1839	1407	0.48	469	100%	0.15	142	29	30.28	6.18
TSC_C03600	420	0.48	TTHA1745	420	0.50	144	96%	0.15	69	19	47.92	13.19
TSC_C21390	729	0.48	TTHA1606	729	0.49	246	98%	0.15	108	30	43.90	12.20
TSC_C13400	486	0.43	TTHA0929	483	0.44	161	99%	0.15	63	11	39.13	6.83
TSC_C20700	2265	0.45	TTHA1383	2274	0.46	765	99%	0.15	398	108	52.03	14.12
TSC_C21840	1392	0.50	TTHA1520	1392	0.52	467	99%	0.15	204	50	43.68	10.71
TSC_C04140	594	0.47	TTHA1819	597	0.48	204	96%	0.15	121	34	59.31	16.67
TSC_C20280	555	0.49	TTHA1490	534	0.51	187	96%	0.15	95	34	50.80	18.18
TSC_C08360	1398	0.46	TTHA0558	1401	0.48	469	99%	0.15	185	45	39.45	9.59
TSC_C04970	828	0.55	TTHA1350	843	0.56	287	96%	0.15	170	47	59.23	16.38
TSC_C24570	882	0.51	TTHA0052	885	0.53	297	98%	0.15	169	41	56.90	13.80
TSC_C20350	687	0.46	TTHA1483	687	0.47	228	100%	0.15	95	25	41.67	10.96
TSC_C15290	279	0.48	TTHA1194	273	0.50	95	95%	0.15	82	28	86.32	29.47
TSC_C13510	912	0.55	TTHA0941	912	0.56	303	100%	0.14	30	7	9.90	2.31
TSC_C05330	351	0.40	TTHA0496	390	0.42	129	94%	0.14	42	4	32.56	3.10

TSC_C08980	1155	0.51	TTHA0690	1155	0.52	385	99%	0.14	184	37	47.79	9.61
TSC_C19550	1149	0.53	TTHA0486	1134	0.54	396	95%	0.14	238	83	60.10	20.96
TSC_C13030	744	0.47	TTHA1145	744	0.48	251	98%	0.14	100	24	39.84	9.56
TSC_C17750	282	0.39	TTHA0177	240	0.40	93	91%	0.14	27	10	29.03	10.75
TSC_C20940	1932	0.50	TTHA1404	1932	0.52	647	99%	0.14	276	56	42.66	8.66
TSC_C22010	1023	0.46	TTHA1570	1038	0.47	345	99%	0.14	100	21	28.99	6.09
TSC_C02790	1176	0.48	TTHA0280	1176	0.49	391	100%	0.14	105	22	26.85	5.63
TSC_C07020	597	0.53	TTHA0530	507	0.54	200	90%	0.14	113	25	56.50	12.50
TSC_C05680	1296	0.45	TTHA0443	1293	0.46	434	99%	0.14	203	70	46.77	16.13
TSC_C18140	2997	0.47	TTHA1251	2994	0.49	1003	99%	0.14	314	60	31.31	5.98
TSC_C17860	1338	0.49	TTHA1296	1320	0.50	447	99%	0.14	182	45	40.72	10.07
TSC_C20300	237	0.47	TTHA1488	237	0.48	78	99%	0.14	38	9	48.72	11.54
TSC_C07650	1413	0.46	TTHA1228	1419	0.48	475	99%	0.14	154	27	32.42	5.68
TSC_C05100	1341	0.42	TTHA1329	1341	0.44	449	99%	0.14	146	27	32.52	6.01
TSC_C20120	978	0.51	TTHA0118	975	0.52	326	99%	0.14	122	14	37.42	4.29
TSC_C03400	501	0.51	TTHA0165	483	0.53	172	94%	0.14	114	36	66.28	20.93
TSC_C00890	1266	0.50	TTHA0277	1248	0.52	431	97%	0.14	231	67	53.60	15.55
TSC_C02020	879	0.50	TTHA0266	777	0.51	304	89%	0.14	136	45	44.74	14.80
TSC_C08420	1557	0.50	TTHA1210	1563	0.51	521	99%	0.14	175	33	33.59	6.33
TSC_C15020	951	0.48	TTHA1157	945	0.50	317	99%	0.13	128	30	40.38	9.46
TSC_C01780	444	0.45	TTHA0247	444	0.46	147	99%	0.13	32	3	21.77	2.04
TSC_C10520	456	0.50	TTHA0965	411	0.51	154	92%	0.13	71	18	46.10	11.69
TSC_C16990	831	0.47	TTHA0857	825	0.48	277	99%	0.13	144	43	51.99	15.52
TSC_C20990	669	0.49	TTHA1409	669	0.50	222	100%	0.13	124	28	55.86	12.61
TSC_C22810	2514	0.50	TTHA0396	2475	0.52	857	97%	0.13	463	148	54.03	17.27
TSC_C00290	858	0.54	TTHA1459	924	0.55	344	84%	0.13	263	113	76.45	32.85
TSC_C24600	1329	0.52	TTHA0056	1341	0.53	450	98%	0.12	229	64	50.89	14.22
TSC_C02300	372	0.39	TTHA0238	360	0.41	126	95%	0.12	58	13	46.03	10.32
TSC_C18750	969	0.50	TTHA1446	969	0.51	322	100%	0.12	110	13	34.16	4.04

TSC_C05970	549	0.44	TTHA0092	549	0.45	183	99%	0.12	51	4	27.87	2.19
TSC_C22410	831	0.49	TTHA1689	831	0.50	276	100%	0.12	77	9	27.90	3.26
TSC_C09500	543	0.49	TTHA0927	504	0.51	202	83%	0.12	147	49	72.77	24.26
TSC_C22190	390	0.42	TTHA1666	390	0.43	129	99%	0.12	40	5	31.01	3.88
TSC_C07470	1545	0.47	TTHA1246	1545	0.48	515	100%	0.12	184	31	35.73	6.02
TSC_C23310	1179	0.48	TTHA1642	1188	0.49	395	99%	0.12	121	11	30.63	2.78
TSC_C23100	465	0.49	TTHA1621	471	0.50	157	98%	0.12	76	29	48.41	18.47
TSC_C06010	546	0.47	TTHA0088	546	0.48	181	99%	0.12	57	8	31.49	4.42
TSC_C03120	375	0.45	TTHA1777	342	0.46	124	95%	0.12	47	15	37.90	12.10
TSC_C06050	360	0.49	TTHA0084	360	0.50	120	98%	0.11	40	8	33.33	6.67
TSC_C21860	1224	0.48	TTHA1524	1224	0.49	408	100%	0.11	133	20	32.60	4.90
TSC_C22360	198	0.43	TTHA1684	219	0.44	72	94%	0.11	33	12	45.83	16.67
TSC_C13960	759	0.50	TTHA0625	759	0.51	252	100%	0.11	124	24	49.21	9.52
TSC_C03330	777	0.45	TTHA0172	786	0.46	265	98%	0.11	118	37	44.53	13.96
TSC_C17190	660	0.49	TTHA0836	657	0.50	221	98%	0.11	111	33	50.23	14.93
TSC_C00260	627	0.54	TTHA0007	645	0.55	219	96%	0.11	144	49	65.75	22.37
TSC_C07160	453	0.42	TTHA0795	453	0.43	151	99%	0.11	64	24	42.38	15.89
TSC_C21490	429	0.48	TTHA1615	429	0.49	143	99%	0.11	52	6	36.36	4.20
TSC_C23060	477	0.49	TTHA0382	477	0.50	160	98%	0.11	101	27	63.13	16.88
TSC_C05480	759	0.55	TTHA1438	714	0.56	252	97%	0.11	117	34	46.43	13.49
TSC_C21610	651	0.46	TTHA1501	618	0.47	219	96%	0.11	116	35	52.97	15.98
TSC_C20790	633	0.49	TTHA1390	576	0.50	217	91%	0.11	121	34	55.76	15.67
TSC_C01840	792	0.47	TTHA1915	792	0.48	264	99%	0.11	107	21	40.53	7.95
TSC_C19570	1056	0.52	TTHA0491	1056	0.53	356	98%	0.11	175	35	49.16	9.83
TSC_C02290	972	0.50	TTHA0239	972	0.51	323	100%	0.11	134	37	41.49	11.46
TSC_C22380	720	0.49	TTHA1686	720	0.50	239	100%	0.10	94	20	39.33	8.37
TSC_C12180	456	0.51	TTHA0969	513	0.52	172	92%	0.10	49	15	28.49	8.72
TSC_C08030	873	0.47	TTHA1262	873	0.48	293	99%	0.10	116	23	39.59	7.85
TSC_C08640	1488	0.49	TTHA1286	1494	0.50	514	96%	0.10	280	72	54.47	14.01

TSC_C06340	1134	0.46	TTHA1343	1134	0.47	377	100%	0.10	150	33	39.79	8.75
TSC_C11840	1332	0.48	TTHA0678	1308	0.49	445	98%	0.10	165	22	37.08	4.94
TSC_C05930	1410	0.50	TTHA0096	1410	0.51	473	99%	0.10	232	59	49.05	12.47
TSC_C18060	435	0.51	TTHA0828	447	0.52	154	94%	0.10	104	33	67.53	21.43
TSC_C17920	780	0.52	TTHA1094	816	0.53	272	97%	0.10	128	29	47.06	10.66
TSC_C01920	897	0.50	TTHA1907	843	0.51	299	96%	0.10	125	36	41.81	12.04
TSC_C24170	771	0.45	TTHA1889	771	0.46	259	98%	0.10	128	37	49.42	14.29
TSC_C08310	357	0.46	TTHA0553	357	0.47	118	99%	0.10	46	11	38.98	9.32
TSC_C05810	1191	0.52	TTHA0411	1194	0.53	402	98%	0.09	199	54	49.50	13.43
TSC_C19750	549	0.43	TTHA0189	546	0.44	222	77%	0.09	149	69	67.12	31.08
TSC_C20490	4479	0.53	TTHA1468	4479	0.54	1496	100%	0.09	545	101	36.43	6.75
TSC_C08870	435	0.52	TTHA0733	435	0.53	152	94%	0.09	115	48	75.66	31.58
TSC_C23180	1554	0.45	TTHA1629	1551	0.46	526	98%	0.09	333	101	63.31	19.20
TSC_C01820	582	0.49	TTHA1917	567	0.50	194	98%	0.09	107	26	55.15	13.40
TSC_C01790	690	0.45	TTHA0246	690	0.46	229	100%	0.09	73	11	31.88	4.80
TSC_C22450	318	0.42	TTHA1693	318	0.43	105	99%	0.09	32	6	30.48	5.71
TSC_C01930	1041	0.46	TTHA1904	1038	0.47	347	99%	0.09	143	28	41.21	8.07
TSC_C02850	1107	0.49	TTHA1108	1107	0.50	370	99%	0.09	198	65	53.51	17.57
TSC_C03580	333	0.39	TTHA1747	333	0.40	111	98%	0.09	39	11	35.14	9.91
TSC_C13480	996	0.51	TTHA0938	981	0.52	331	99%	0.09	168	22	50.76	6.65
TSC_C21730	660	0.52	TTHA1509	660	0.53	223	98%	0.09	121	38	54.26	17.04
TSC_C08100	1521	0.45	TTHA0543	1521	0.46	506	100%	0.09	161	23	31.82	4.55
TSC_C18280	747	0.50	TTHA0785	738	0.51	253	97%	0.09	159	43	62.85	17.00
TSC_C09640	2145	0.50	TTHA1139	2142	0.51	720	99%	0.08	276	54	38.33	7.50
TSC_C02060	1536	0.50	TTHA0270	1533	0.51	514	99%	0.08	229	63	44.55	12.26
TSC_C03650	552	0.50	TTHA1742	552	0.51	185	98%	0.08	79	13	42.70	7.03
TSC_C02910	1323	0.47	TTHA0649	1266	0.48	506	82%	0.08	410	140	81.03	27.67
TSC_C24690	2652	0.48	TTHA0161	2637	0.49	889	99%	0.08	311	76	34.98	8.55
TSC_C05300	627	0.47	TTHA1312	627	0.48	208	100%	0.08	83	17	39.90	8.17

TSC_C10000	543	0.53	TTHA0631	567	0.54	188	97%	0.08	87	10	46.28	5.32
TSC_C02330	651	0.53	TTHA0235	651	0.54	221	97%	0.08	90	23	40.72	10.41
TSC_C01970	846	0.48	TTHA1900	846	0.49	284	99%	0.08	107	27	37.68	9.51
TSC_C20090	1434	0.48	TTHA0115	1434	0.49	479	99%	0.08	195	43	40.71	8.98
TSC_C14970	738	0.52	TTHA1164	765	0.53	259	96%	0.08	143	37	55.21	14.29
TSC_C06810	972	0.52	TTHA0569	1032	0.53	395	81%	0.07	335	137	84.81	34.68
TSC_C14400	216	0.38	TTHA1073	216	0.39	71	99%	0.07	28	4	39.44	5.63
TSC_C20550	777	0.50	TTHA1370	762	0.51	273	93%	0.07	163	56	59.71	20.51
TSC_C19690	1020	0.47	TTHA0472	1023	0.48	340	100%	0.07	127	25	37.35	7.35
TSC_C00160	540	0.43	TTHA0194	543	0.43	184	97%	0.07	116	33	63.04	17.93
TSC_C20380	420	0.47	TTHA1478	459	0.47	156	92%	0.07	70	18	44.87	11.54
TSC_C17170	723	0.45	TTHA0838	690	0.46	241	97%	0.07	80	22	33.20	9.13
TSC_C04930	444	0.45	TTHA1354	450	0.46	149	99%	0.07	68	12	45.64	8.05
TSC_C04450	345	0.49	TTHA0426	360	0.49	123	94%	0.07	76	26	61.79	21.14
TSC_C00370	750	0.52	TTHA0360	492	0.52	251	78%	0.07	80	13	31.87	5.18
TSC_C22470	2076	0.47	TTHA1695	2076	0.47	692	100%	0.07	168	19	24.28	2.75
TSC_C06890	387	0.44	TTHA0524	387	0.45	129	98%	0.06	54	11	41.86	8.53
TSC_C04380	975	0.37	TTHA0034	966	0.38	324	99%	0.06	21	8	6.48	2.47
TSC_C04860	363	0.50	TTHA1361	366	0.51	128	93%	0.06	101	34	78.91	26.56
TSC_C19780	192	0.34	TTHA0462	192	0.35	63	98%	0.06	33	10	52.38	15.87
TSC_C23220	1011	0.46	TTHA1633	1011	0.46	337	99%	0.06	131	20	38.87	5.93
TSC_C17210	849	0.51	TTHA0603	849	0.51	284	99%	0.06	169	51	59.51	17.96
TSC_C12150	972	0.45	TTHA0972	972	0.46	323	100%	0.06	90	14	27.86	4.33
TSC_C08550	369	0.42	TTHA0719	369	0.43	122	99%	0.06	44	6	36.07	4.92
TSC_C19300	621	0.46	TTHA1589	621	0.46	209	98%	0.06	107	20	51.20	9.57
TSC_C22040	522	0.51	TTHA1575	528	0.51	178	97%	0.06	71	18	39.89	10.11
TSC_C23240	381	0.46	TTHA1635	381	0.46	127	98%	0.06	29	3	22.83	2.36
TSC_C03160	1146	0.46	TTHA1774	1107	0.47	381	98%	0.06	128	38	33.60	9.97
TSC_C00420	552	0.49	TTHA1791	558	0.49	185	99%	0.06	62	16	33.51	8.65

TSC_C03080	1254	0.54	TTHA1781	1251	0.54	426	97%	0.06	218	56	51.17	13.15
TSC_C18150	972	0.56	TTHA1252	981	0.57	331	98%	0.06	179	51	54.08	15.41
TSC_C15570	1284	0.46	TTHA1065	1281	0.47	427	100%	0.06	146	21	34.19	4.92
TSC_C08350	615	0.42	TTHA0557	615	0.43	204	100%	0.06	62	12	30.39	5.88
TSC_C22290	543	0.50	TTHA1677	543	0.51	181	99%	0.06	74	17	40.88	9.39
TSC_C22200	381	0.48	TTHA1667	381	0.48	126	99%	0.06	37	2	29.37	1.59
TSC_C05290	1089	0.47	TTHA1313	1101	0.48	370	98%	0.05	162	32	43.78	8.65
TSC_C07220	780	0.48	TTHA0801	774	0.49	270	95%	0.05	156	36	57.78	13.33
TSC_C13600	753	0.50	TTHA0947	753	0.51	250	100%	0.05	123	28	49.20	11.20
TSC_C20150	705	0.52	TTHA0461	705	0.52	236	99%	0.05	91	19	38.56	8.05
TSC_C08530	684	0.52	TTHA1200	684	0.52	227	100%	0.05	105	21	46.26	9.25
TSC_C13440	1266	0.47	TTHA0760	1266	0.48	421	100%	0.05	160	27	38.00	6.41
TSC_C03920	753	0.48	TTHA1838	753	0.49	250	100%	0.05	80	17	32.00	6.80
TSC_C01940	810	0.47	TTHA1903	810	0.48	270	99%	0.05	112	28	41.48	10.37
TSC_C00460	810	0.52	TTHA0069	882	0.52	300	93%	0.04	191	55	63.67	18.33
TSC_C23230	1881	0.45	TTHA1634	1869	0.45	628	99%	0.04	198	56	31.53	8.92
TSC_C17350	654	0.47	TTHA1502	684	0.48	252	86%	0.04	186	79	73.81	31.35
TSC_C04370	1131	0.46	TTHA0035	1104	0.46	376	99%	0.04	16	6	4.26	1.60
TSC_C00360	207	0.39	TTHA0359	207	0.40	68	99%	0.04	33	10	48.53	14.71
TSC_C08450	777	0.44	TTHA1206	777	0.44	259	99%	0.04	118	23	45.56	8.88
TSC_C02760	540	0.50	TTHA0282	534	0.51	179	99%	0.04	79	20	44.13	11.17
TSC_C02240	267	0.46	TTHA0243	267	0.46	88	99%	0.04	24	5	27.27	5.68
TSC_C13630	606	0.48	TTHA0904	573	0.49	203	96%	0.04	108	25	53.20	12.32
TSC_C09570	1449	0.52	TTHA0283	1389	0.53	576	78%	0.04	428	159	74.31	27.60
TSC_C22350	318	0.43	TTHA1683	318	0.44	105	99%	0.03	31	7	29.52	6.67
TSC_C24590	1560	0.50	TTHA0054	1536	0.51	526	98%	0.03	259	62	49.24	11.79
TSC_C09260	1656	0.47	TTHA1039	1656	0.47	551	100%	0.03	127	14	23.05	2.54
TSC_C04230	3360	0.51	TTHA1813	3360	0.52	1121	100%	0.03	349	73	31.13	6.51
TSC_C06040	546	0.47	TTHA0085	546	0.47	181	99%	0.03	47	2	25.97	1.10

TSC_C14820	1110	0.43	TTHA0685	759	0.43	370	81%	0.03	128	35	34.59	9.46
TSC_C12170	759	0.49	TTHA0970	747	0.49	256	97%	0.03	122	38	47.66	14.84
TSC_C01910	165	0.41	TTHA1908	165	0.41	54	98%	0.02	23	3	42.59	5.56
TSC_C15180	360	0.43	TTHA1184	360	0.43	122	97%	0.02	102	33	83.61	27.05
TSC_C18970	1734	0.49	TTHA1454	1734	0.50	579	99%	0.02	149	19	25.73	3.28
TSC_C09620	462	0.48	TTHA0295	435	0.49	169	86%	0.02	122	47	72.19	27.81
TSC_C01390	876	0.43	TTHA0355	876	0.43	291	100%	0.02	139	30	47.77	10.31
TSC_C24990	495	0.52	TTHA1966	627	0.52	224	79%	0.01	110	39	49.11	17.41
TSC_C07290	591	0.53	TTHA0808	534	0.53	204	90%	0.01	114	34	55.88	16.67
TSC_C06590	924	0.49	TTHA0142	1137	0.49	380	89%	0.01	132	41	34.74	10.79
TSC_C07680	1014	0.45	TTHA1227	1014	0.45	339	99%	0.01	126	28	37.17	8.26
TSC_C21830	2208	0.51	TTHA1518	2178	0.51	742	98%	0.01	321	76	43.26	10.24
TSC_C22130	525	0.49	TTHA1660	513	0.49	176	97%	0.01	98	32	55.68	18.18
TSC_C03760	654	0.46	TTHA1731	636	0.47	224	95%	0.01	103	22	45.98	9.82
TSC_C08810	2706	0.49	TTHA0726	2709	0.50	906	99%	0.01	296	69	32.67	7.62
TSC_C02000	699	0.52	TTHA1897	699	0.52	237	97%	0.00	142	45	59.92	18.99
TSC_C05950	288	0.48	TTHA0094	288	0.49	95	99%	0.00	30	8	31.58	8.42
TSC_C02600	1290	0.53	TTHA1005	1314	0.53	463	93%	0.00	358	128	77.32	27.65
TSC_C14740	738	0.52	TTHA0750	738	0.52	248	98%	0.00	110	20	44.35	8.06
TSC_C10470	957	0.47	TTHA0962	960	0.47	328	97%	0.00	149	43	45.43	13.11
TSC_C03730	2025	0.52	TTHA1733	2076	0.52	714	95%	0.00	466	143	65.27	20.03
TSC_C10290	1035	0.37	TTHA0948	1038	0.37	354	97%	0.00	128	36	36.16	10.17
TSC_C05520	822	0.51	TTHA1434	816	0.51	274	99%	0.00	118	26	43.07	9.49
TSC_C15030	777	0.45	TTHA1170	786	0.45	266	97%	0.00	144	29	54.14	10.90
TSC_C05370	318	0.46	TTHA0499	318	0.46	105	99%	-0.01	49	16	46.67	15.24
TSC_C17040	786	0.55	TTHA0852	786	0.55	263	99%	-0.01	133	36	50.57	13.69
TSC_C12960	1473	0.51	TTHA0089	1317	0.50	576	76%	-0.01	378	159	65.63	27.60
TSC_C07000	489	0.46	TTHA0528	513	0.45	171	96%	-0.01	89	21	52.05	12.28
TSC_C22880	744	0.45	TTHA0387	741	0.45	250	98%	-0.01	98	24	39.20	9.60

TSC_C02770	264	0.46	TTHA0281	264	0.46	87	99%	-0.02	30	6	34.48	6.90
TSC_C00270	663	0.45	TTHA0008	663	0.45	226	97%	-0.02	112	29	49.56	12.83
TSC_C07170	483	0.45	TTHA0796	468	0.45	179	86%	-0.02	158	69	88.27	38.55
TSC_C08790	1437	0.45	TTHA1272	1437	0.45	479	100%	-0.02	105	6	21.92	1.25
TSC_C24200	2001	0.50	TTHA1892	1998	0.50	672	99%	-0.02	244	49	36.31	7.29
TSC_C13940	537	0.48	TTHA1057	543	0.48	181	98%	-0.02	66	12	36.46	6.63
TSC_C15490	1083	0.44	TTHA0766	1086	0.44	361	100%	-0.03	251	58	69.53	16.07
TSC_C24210	2664	0.47	TTHA1893	2787	0.47	1050	84%	-0.03	638	272	60.76	25.90
TSC_C03940	1317	0.48	TTHA1836	1308	0.48	439	99%	-0.03	133	26	30.30	5.92
TSC_C22370	426	0.45	TTHA1685	426	0.45	141	99%	-0.03	41	7	29.08	4.96
TSC_C07500	804	0.47	TTHA1242	804	0.47	267	100%	-0.03	125	28	46.82	10.49
TSC_C00170	1500	0.47	TTHA0195	1500	0.46	502	99%	-0.03	253	57	50.40	11.35
TSC_C03470	516	0.50	TTHA1761	465	0.50	175	92%	-0.03	82	34	46.86	19.43
TSC_C03780	810	0.50	TTHA1727	816	0.49	307	86%	-0.03	234	79	76.22	25.73
TSC_C06030	621	0.47	TTHA0086	624	0.46	207	99%	-0.04	83	23	40.10	11.11
TSC_C17220	1047	0.51	TTHA0601	1026	0.51	352	98%	-0.04	174	55	49.43	15.63
TSC_C10230	1299	0.52	TTHA0862	1299	0.51	435	99%	-0.04	227	63	52.18	14.48
TSC_C13070	672	0.47	TTHA1148	660	0.47	230	95%	-0.04	119	31	51.74	13.48
TSC_C17120	3192	0.45	TTHA0843	2007	0.45	1065	77%	-0.05	222	48	20.85	4.51
TSC_C24510	219	0.43	TTHA1669	213	0.42	78	89%	-0.05	57	18	73.08	23.08
TSC_C13710	681	0.52	TTHA0896	687	0.51	230	98%	-0.05	103	21	44.78	9.13
TSC_C02550	789	0.59	TTHA1001	507	0.59	265	76%	-0.06	110	33	41.51	12.45
TSC_C02650	765	0.54	TTHA0290	765	0.53	254	100%	-0.06	127	32	50.00	12.60
TSC_C20860	711	0.52	TTHA1395	741	0.51	253	95%	-0.06	134	43	52.96	17.00
TSC_C03570	378	0.45	TTHA1748	378	0.45	126	98%	-0.07	50	10	39.68	7.94
TSC_C03450	1353	0.44	TTHA1763	1347	0.44	453	99%	-0.07	162	46	35.76	10.15
TSC_C21370	267	0.40	TTHA1604	267	0.39	88	99%	-0.07	33	10	37.50	11.36
TSC_C22440	621	0.48	TTHA1692	621	0.48	206	100%	-0.07	83	15	40.29	7.28
TSC_C15530	2055	0.54	TTHA0706	2055	0.53	684	100%	-0.07	77	30	11.26	4.39

TSC_C21590	918	0.52	TTHA1499	891	0.51	306	98%	-0.07	128	36	41.83	11.76
TSC_C09110	1008	0.55	TTHA0416	969	0.54	409	75%	-0.07	265	118	64.79	28.85
TSC_C25010	309	0.45	TTHA1967	228	0.44	105	80%	-0.07	46	11	43.81	10.48
TSC_C08590	792	0.50	TTHA0723	798	0.49	266	99%	-0.07	135	36	50.75	13.53
TSC_C18810	1194	0.43	TTHA0377	900	0.42	397	86%	-0.08	126	24	31.74	6.05
TSC_C05340	399	0.46	TTHA1360	354	0.45	147	81%	-0.08	119	49	80.95	33.33
TSC_C08540	618	0.45	TTHA0718	618	0.44	205	100%	-0.08	70	13	34.15	6.34
TSC_C08280	474	0.47	TTHA0551	435	0.46	157	95%	-0.08	45	14	28.66	8.92
TSC_C05790	960	0.49	TTHA0409	894	0.48	327	94%	-0.09	155	39	47.40	11.93
TSC_C04510	1245	0.46	TTHA0022	1245	0.45	416	99%	-0.09	137	24	32.93	5.77
TSC_C08470	786	0.47	TTHA1204	774	0.46	269	96%	-0.09	134	28	49.81	10.41
TSC_C03870	285	0.48	TTHA1719	285	0.47	95	98%	-0.09	37	6	38.95	6.32
TSC_C09650	270	0.43	TTHA1138	270	0.42	89	99%	-0.09	42	10	47.19	11.24
TSC_C03880	201	0.41	TTHA1718	201	0.40	67	97%	-0.09	35	11	52.24	16.42
TSC_C08520	603	0.51	TTHA1203	600	0.50	206	96%	-0.10	111	26	53.88	12.62
TSC_C03680	531	0.49	TTHA1739	537	0.48	189	93%	-0.10	87	17	46.03	8.99
TSC_C07240	813	0.49	TTHA0803	816	0.48	281	96%	-0.10	151	53	53.74	18.86
TSC_C02320	1713	0.50	TTHA0236	1332	0.49	576	86%	-0.10	210	51	36.46	8.85
TSC_C01620	615	0.50	TTHA0333	663	0.49	227	92%	-0.11	125	39	55.07	17.18
TSC_C20410	834	0.41	TTHA1475	828	0.40	279	99%	-0.11	116	21	41.58	7.53
TSC_C03810	516	0.44	TTHA1725	369	0.43	176	79%	-0.11	76	19	43.18	10.80
TSC_C22480	471	0.47	TTHA1696	471	0.46	156	99%	-0.11	53	6	33.97	3.85
TSC_C21510	324	0.45	TTHA1617	333	0.44	111	97%	-0.11	43	8	38.74	7.21
TSC_C08240	306	0.51	TTHA0547	312	0.50	105	96%	-0.12	58	23	55.24	21.90
TSC_C20590	759	0.46	TTHA1373	759	0.44	254	99%	-0.12	102	25	40.16	9.84
TSC_C22890	276	0.44	TTHA0386	222	0.43	92	87%	-0.13	27	7	29.35	7.61
TSC_C17730	696	0.49	TTHA1453	699	0.47	239	96%	-0.13	154	61	64.44	25.52
TSC_C13700	249	0.41	TTHA0897	249	0.40	82	99%	-0.13	42	15	51.22	18.29
TSC_C22610	669	0.44	TTHA0221	669	0.43	223	99%	-0.13	96	26	43.05	11.66

TSC_C06650	1176	0.55	TTHA1027	1176	0.53	392	99%	-0.14	54	16	13.78	4.08
TSC_C02740	1203	0.49	TTHA0284	1203	0.48	402	99%	-0.15	131	24	32.59	5.97
TSC_C22260	183	0.40	TTHA1674	183	0.39	60	98%	-0.15	33	12	55.00	20.00
TSC_C14780	855	0.48	TTHA0746	855	0.47	284	100%	-0.15	134	35	47.18	12.32
TSC_C03290	222	0.42	TTHA0175	222	0.40	73	99%	-0.16	20	3	27.40	4.11
TSC_C06500	462	0.48	TTHA0134	462	0.47	155	98%	-0.16	67	13	43.23	8.39
TSC_C19600	339	0.44	TTHA0494	339	0.42	118	94%	-0.17	86	23	72.88	19.49
TSC_C07190	690	0.53	TTHA0159	654	0.51	241	92%	-0.17	132	45	54.77	18.67
TSC_C14280	381	0.46	TTHA0627	375	0.44	131	94%	-0.17	87	32	66.41	24.43
TSC_C05840	243	0.39	TTHA0414	243	0.37	80	99%	-0.18	20	4	25.00	5.00
TSC_C10770	825	0.49	TTHA1338	1020	0.48	357	83%	-0.18	267	92	74.79	25.77
TSC_C23940	1746	0.53	TTHA0771	1680	0.51	694	78%	-0.19	516	193	74.35	27.81
TSC_C22330	333	0.44	TTHA1681	333	0.42	110	99%	-0.19	44	14	40.00	12.73
TSC_C06250	303	0.40	TTHA0516	303	0.39	100	99%	-0.19	50	10	50.00	10.00
TSC_C02920	1122	0.43	TTHA0886	1119	0.41	376	99%	-0.19	213	61	56.65	16.22
TSC_C21440	228	0.36	TTHA1610	234	0.34	79	95%	-0.19	20	5	25.32	6.33
TSC_C08700	210	0.49	TTHA1281	210	0.47	69	99%	-0.20	37	12	53.62	17.39
TSC_C05850	738	0.54	TTHA0415	738	0.52	247	99%	-0.21	94	18	38.06	7.29
TSC_C12640	1323	0.46	TTHA0742	1233	0.44	463	91%	-0.22	299	84	64.58	18.14
TSC_C02500	1056	0.53	TTHA1715	1068	0.51	373	94%	-0.23	204	70	54.69	18.77
TSC_C18130	735	0.48	TTHA0821	735	0.46	244	100%	-0.23	99	23	40.57	9.43
TSC_C07960	990	0.50	TTHA1268	645	0.47	334	77%	-0.25	124	39	37.13	11.68
TSC_C22730	387	0.49	TTHA1585	384	0.47	130	97%	-0.25	73	14	56.15	10.77
TSC_C13850	462	0.43	TTHA1031	441	0.41	157	94%	-0.26	70	23	44.59	14.65
TSC_C03700	204	0.46	TTHA1737	207	0.44	71	93%	-0.26	32	12	45.07	16.90
TSC_C21750	579	0.49	TTHA1510	582	0.46	197	97%	-0.26	114	35	57.87	17.77
TSC_C06740	279	0.48	TTHA1021	228	0.45	96	84%	-0.28	67	21	69.79	21.88
TSC_C14110	330	0.48	TTHA0611	330	0.45	110	98%	-0.30	48	13	43.64	11.82
TSC_C03430	636	0.48	TTHA1765	639	0.44	217	97%	-0.31	107	32	49.31	14.75

TSC_C17320	1266	0.57	TTHA0975	1047	0.54	466	79%	-0.31	285	123	61.16	26.39
TSC_C02570	1203	0.58	TTHA1003	1218	0.55	421	95%	-0.33	246	75	58.43	17.81
TSC_C05980	1119	0.44	TTHA0091	1098	0.41	375	98%	-0.33	145	35	38.67	9.33
TSC_C06970	597	0.54	TTHA1352	648	0.50	252	78%	-0.33	148	69	58.73	27.38
TSC_C06720	2049	0.54	TTHA1054	2505	0.51	946	75%	-0.33	636	256	67.23	27.06
TSC_C03840	213	0.46	TTHA1721	213	0.43	71	97%	-0.33	25	7	35.21	9.86
TSC_C22520	429	0.48	TTHA0212	432	0.44	149	95%	-0.34	91	32	61.07	21.48
TSC_C08600	1332	0.47	TTHA0724	1305	0.43	456	96%	-0.35	367	108	80.48	23.68
TSC_C00430	321	0.48	TTHA1792	330	0.44	109	98%	-0.35	61	14	55.96	12.84
TSC_C09090	774	0.54	TTHA1738	711	0.50	305	76%	-0.37	203	83	66.56	27.21
TSC_C05360	393	0.46	TTHA0498	399	0.42	134	97%	-0.38	85	24	63.43	17.91
TSC_C13570	1464	0.63	TTHA0029	1479	0.59	562	85%	-0.39	291	109	51.78	19.40
TSC_C09100	1479	0.50	TTHA1463	1683	0.46	627	81%	-0.40	434	183	69.22	29.19
TSC_C02820	939	0.52	TTHA0278	936	0.48	317	98%	-0.40	163	48	51.42	15.14
TSC_C24700	783	0.47	TTHA1929	798	0.42	277	94%	-0.44	179	55	64.62	19.86
TSC_C05800	705	0.50	TTHA0410	642	0.45	252	87%	-0.50	189	64	75.00	25.40
TSC_C19180	1539	0.57	TTHA1355	1533	0.51	523	97%	-0.68	279	91	53.35	17.40
TSC_C14730	1167	0.51	TTHA0751	1167	0.38	400	97%	-1.27	272	86	68.00	21.50
Correlation coefficient[‡]:											0.30	0.38

NOTES:

*Top 100 MFE values are highlighted by grey cell background.

[†]Frequencies of nucleotide substitutions per 300 nucleotides and amino acid substitutions per 100 amino acids

[‡]Correlation coefficients between Δ MFE values and normalized frequencies of nucleotide and amino acid substitutions

6.5 Appendix D

6.5.1 Distribution of amino acid residue and property analysis

Table 6.5.1: *p*-values for *t*-test and Wilcoxon *t*-test (*) analysis of distribution of amino acid residues and properties of orthologous sequences of *T. scotoductus* SA-01 and *T. thermophilus* HB27.

Amino acid residue	Property Category	p-value	Amino acid property	p-value
Ala	Tiny, small, hydrophobic	5.28e-16	Acid	2.2e-16
*Arg	Polar, positively charged	2.2e-16	*Aromatic	0.07219
Asn	Small, polar	0.004904	Aliphatic	0.3708
*Asp	Small, polar, negatively charged	0.9817	Basic	0.01576
Cys	Hydrophobic, small	0.7697	Charged	0.1786
Glu	Polar, negatively charged	0.4175	Positively charged	0.01576
Gly	Tiny, small, hydrophobic	0.003554	Hydrophobic	0.657
Gln	polar	9.36e-12	Polar	1.85e-05
His	Polar, positively charged, aromatic	0.135	Sulphur	0.008966
Ile	Aliphatic, hydrophobic	1.309e-09	Tiny	2.2e-16
Leu	Aliphatic, hydrophobic	0.5715	*Small	0.002373
Lys	Hydrophobic, positively charged	0.0001872		
Met	hydrophobic	0.0003932		
Phe	Hydrophobic, aromatic	0.8846		
Pro	Small	0.006934		
Ser	Tiny, small, polar	2.4e-14		
Tyr	Hydrophobic, polar, aromatic	0.3892		
Thr	Hydrophobic, small, polar	8.205e-06		
Trp	Hydrophobic, aromatic, polar	0.1069		
*Val	Hydrophobic, aliphatic, small	0.001578		