# POTENTIAL ASSOCIATION BETWEEN TSETSE FLY AND ITS ENDOSYMBIONTS IN B-VITAMINS BIOSYNTHESIS

**KEVIN KAMANYI MARUCHA**

**SM14/3050/11**

**A Thesis Submitted to the Graduate School in Partial Fulfilment for the Requirements of the award of the Master of Science Degree in Biochemistry of Egerton University**

**EGERTON UNIVERSITY**

**AUGUST, 2014**

# DECLARATION AND RECOMMENDATION

**DECLARATION**

I hereby declare that this research thesis is my original work and has not been submitted wholly or in part to any institution for award of any degree.

Kevin Kamanyi Marucha
Reg No. SM14/3050/11

Signature _____          Date _____

**RECOMMENDATION**

This thesis has been submitted with our approval as supervisors for examination according to Egerton University regulations.

Dr. Vincent A. Owino
Egerton University

Signature _____          Date _____

Dr. Paul Mireji
Yale University, School of Public Health

Signature _____          Date _____

# DEDICATION

To Daniel Marucha

# ACKNOWLEDGEMENT

# ABSTRACT

Tsetse flies are the cyclical vectors of African trypanosomes, flagellated protozoa parasites that cause sleeping sickness in man and nagana in cattle, diseases collectively known as African trypanosomiasis. Vector reduction and chemotherapy are the main strategies for controlling trypanosomiasis but their limitations necessitate improvement of current and/or development of novel methods. Tsetse has three maternally transmitted bacterial endosymbionts namely *Wigglesworthia*, *Sodalis* and *Wolbachia* which are potential targets in developing paratransgenesis control approaches. *Wigglesworthia* is thought to provide B-vitamins to its host, but its B-vitamins biosynthesis pathways are incomplete while those for tsetse remain unknown. Using the recently published *Glossina morsitans morsitans* genome, bioinformatics approach was applied to annotate tsetse B-vitamins and cofactors biosynthesis enzymes. Homology based searches, protein domain architecture interrogations, comparative and RNA seq analyses, polymerase chain reaction (PCR) amplification and sequencing were performed for annotation and interrogation of possible tsetse-endosymbiont interaction. Thirty four genes encode 26 enzymes involved in biosynthesis of B-vitamins namely thiamine, riboflavin, niacin, folate, pantothenate, pyridoxine and their resultant cofactors. Vitamins biosynthesis pathways are incomplete, a phenomenon also observed in insects and humans. However, cofactors biosynthesis pathways are complete in all the organisms analyzed. Notably, integration of tsetse-endosymbionts systems gives complete or near complete pathways suggesting potential association. Moreover, expression of these enzymes depends on availability of endosymbionts, and presence of vitamin transporters in both organisms support association at biochemical level. This association can be through integrated biosynthetic pathways that involve import/export transporters of metabolic intermediates and sharing of final products for cofactor biosynthesis. Therefore, one of the possible bases of the symbiotic association of tsetse and its endosymbinots is B-vitamins biosynthesis.

**TABLE OF CONTENTS**

# LIST OF TABLES

# LIST OF FIGURES

## LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| AAT | Animal African trypanosomiasis |
| BLAST | Basic local alignment search tool |
| CoA | Coenzyme A |
| DDT | Dichlorodiphenyltrichloroethane |
| DNA | Deoxyribonucleic acid |
| EDTA | Ethylenediaminetetraacetic acid |
| FAD | Flavin adenosine dinucleotide |
| FMN | Flavin mononucleotide |
| HAT | Human African trypanosomiasis |
| ICIPE | International centre of insect physiology and ecology |
| IPTG | Isopropyl $\beta$-D-1-thiogalactopyranoside |
| LB | Luria-Bertani |
| Mbp | Mega base pairs |
| NAD | Nicotinamide adenosine dinucleotide |
| NCBI | National center for biotechnology information |
| NGS | Next generation sequencing |
| ORF | Open reading frame |
| PCR | Polymerase chain reaction |
| PLP | Pyridoxal phosphate |
| RNA | Ribonucleic acid |
| RPKM | reads per kilobase pair per million reads |
| rpm | Revolutions per minute |
| SDS | Sodium dodecyl sulphate |
| THF | Tetrahydrofolate |
| TPP | Thiamine pyrophosphate |
| X-Gal | 5-bromo-4-chloro-3-indolyl-$\beta$-D-galactopyranoside |

# CHAPTER ONE

# INTRODUCTION

## 1.1 Background information

Tsetse flies (Diptera: *Glossinidae*) are vectors of African trypanosomes, flagellated parasites that cause diseases termed African trypanosomiasis. The disease is called sleeping sickness (human African trypanosomiasis, HAT) in man and Nagana (animal African trypanosomiasis, AAT) in cattle. The vector is found in 37 countries in Africa, where approximately 70 million people and 60 million cattle are at risk of the disease (FAO, 2013). About 30,000 actual cases of HAT are reported annually with a disability adjusted life years (DALY) of 1.35 million (Fèvre *et al.*, 2008; Simarro *et al.*, 2011). However, recent data reports only 7,139 new cases (Simarro *et al.*, 2011) which is a decline from previous numbers. This is possibly an underestimation due to geopolitical instability, technical, infrastructural deficiencies in some affected regions and limitation in diagnosis (Odiit *et al.*, 2005; Berrang, 2007). Nagana on the other hand causes over 3 million cattle deaths annually (FAO, 2013). Further, the disease reduces milk and meat production, animal for draught power and manure (Shaw, 2009). With the disease affecting both man and his livestock, it prevents both crop and livestock keeping resulting into total agricultural loss of approximately 4.75 billion US dollars annually (FAO, 2013). Despite such huge impact, the disease is among the most neglected tropical disease (NTDs), with the available drugs limited by chemotoxicity and resistance (Anene *et al.*, 2001; Kennedy, 2006). Therefore, there is an urgent need for improvement of current and/or development of novel control and diagnostic strategies.

The main control strategies for African trypanosomiasis are chemotherapy and blocking of parasite transmission through tsetse fly vector control. In chemotherapy, the drugs applied include pentamidine, suramin, melasoprol and eflornithine, and have been in use since 1950s and currently suffer from resistance (Balasegaram *et al.*, 2009). In addition, melasoprol is cytotoxic, causing 5% deaths in patients (Kennedy, 2006). No new drug is in the pipeline because it is an unattractive market for pharmaceuticals due to poor economic ability of the affected people. Vaccine development is not feasible due to the systematic and periodic change of surface coat by the parasite, a phenomenon termed antigenic variation.

Various vector control approaches have been applied with some success. These include sterile insect technology (SIT), traps and targets, insecticides and the push-pull method. In

1995 *Glossina austeni* was successfully eliminated from Unguja island of Zanzibar using sterile male tsetse flies (Vreysen *et al.*, 2000). However, this approach requires a geographically isolated population to prevent reinvasion of cleared lands, hence has limited application in mainland (Politi *et al.*, 1995). Use of insecticides such as dichlorodiphenyltrichloroethane (DDT) is labour intensive, affects non-target insects and is not environmentally friendly. In addition, resistance to insecticides have rendered methods like push-pull and targets ineffective. In general, all these methods require community participation, awareness and resources, which are seldom assured (Politi *et al.*, 1995). Together, there is urgent need for improvement of current and/or development of novel control strategies.

Tsetse flies harbour bacterial symbionts, a feature that can be exploited for development of novel symbiont-based control approaches (Aksoy, 2003). These symbionts are bacteria that have coevolved with the host to benefit each other. They include primary obligate symbiont of genus *Wigglesworthia* (Aksoy, 1995), secondary symbionts of genus *Sodalis* (Dale and Maudlin, 1999) and *Wolbachia*, which is not always found in all tsetse fly species (Doudoumis *et al.*, 2012). Recent demonstration of importance of some of the endosymbiont in tsetse fly survival (Pais *et al.*, 2008) and possibilities of generating transgenic endosymbionts expressing trypanocidal factors (Medlock *et al.*, 2013) makes symbiont based vector control methods worth investigating further. Notably, transgenic *Sodalis glossinidius* can be used as a paratransgenesis model because of its ability to grow in culture (Beard *et al.*, 1993). It can then be propagated into the tsetse population using the cytoplasmic incompatibility of *Wolbachia* (Medlock *et al.*, 2013). This technique is promising but has not been implemented due to the challenge of obtaining paratransformed progeny of the fly and maintaining the paratransformed flies alive for long. In addition, insight into tsetse-endosymbiont interactions at molecular and biochemical level remains unknown or limited. Improved knowledge on these interactions will permit development of endosymbionts as stable control method.

Tsetse flies are entirely blood feeders - hematophagous. Their restricted diet indicates alternative sources of components deficient in the blood such as vitamins (Edwards *et al.*, 1957). Vitamins are the building blocks of many cofactors, which are essential components of enzymes. In tsetse, vitamins and symbionts have been shown to have a role in fecundity, growth and digestion (Nogge, 1976). In another study, antibiotic treated tsetse flies become aposymbiotic and experience reduced fecundity, growth rate, immunity and digestion of

blood meal (Pais *et al.*, 2008). However, a blood meal supplemented with B vitamins (B1, B5, B6, folic acid and biotin) partially restores the fly's ability to reproduce indicating endosymbiont role in vitamin supply and/or metabolism (Pais *et al.*, 2008) hence a potential interaction. However, the molecular and biochemical aspect of this relationship remains poorly understood, and is the focus of this study.

Recent symbionts genome projects (Akman *et al.*, 2002; Toh *et al.*, 2006; Rita *et al.*, 2012) have shown massive reduction in genes coding for enzymes involved in biosynthesis of vitamins and cofactors in endosymbionts. Various studies clearly demonstrate the importance of endosymbiont in survival of tsetse fly and other insects, with the gene reduction indicating high possibility of integration of insect-endosymbiont metabolic processes (Nogge, 1976; Shigenobu *et al.*, 2000; Gil *et al.*, 2003; van Ham *et al.*, 2003; Pais *et al.*, 2008). Snyder *et al.* (2010) suggested that *Wigglesworthia* synthesizes thiamine used by both tsetse fly and *Sodalis*. However, this remains speculative since vitamin metabolism genes in tsetse fly are unknown and some vitamins and cofactors biosynthesis pathways of endosymbionts lack essential enzymes as compared to the free-living bacteria *E. coli*. Therefore, using all possible orthologs from bacteria, plants and animals, and the recently available *Glossina morsitans morsitans* genome sequence, bioinformatics tools were applied in identification and annotation of B-vitamin and cofactor metabolism factors. In addition, tsetse-endosymbiont interaction in vitamin and cofactor metabolism was determined by applying transcriptomics techniques including RNA seq. This analysis suggests involvement of tsetse fly in B-vitamins and cofactors biosynthesis pathways by encoding biosynthesis enzymes including those missing in endosymbionts. In addition, general upregulation of these biosynthesis enzymes in the absence of the obligate endosymbiont *Wigglesworthia* collectively indicate potential molecular and biochemical interaction between tsetse fly and endosymbionts.

## 1.2 Statement of the problem

The knowledge in vector-endosymbiont interaction is poor, despite the potential application of this interaction in development of novel control strategies for African trypanosomiasis. This necessitates investigation of tsetse fly systems involved in vitamin and cofactor biosynthesis as well as the role of endosymbionts in metabolism and/or provision of these biomolecules. Currently the molecular and biochemical aspects of this association remain speculative since fly metabolic processes remain unknown and therefore the need to

investigate it by utilizing the available *Glossina morsitans morsitans* and endosymbiont genomes.

**1.3 Objectives**

**1.3.1 General objective**

To determine the association between *Glossina morsitans morsitans* and its endosymbionts in vitamins and cofactors biosynthesis.

**1.3.2 Specific objectives**

1. To determine the genes that code for the enzymes involved in biosynthesis of vitamins and cofactors in *G. m. morsitans*.
2. To determine the role of endosymbionts in expression of genes involved in vitamins and cofactors biosynthesis in *G. m. morsitans*.

**1.4 Hypotheses**

1. There are no genes that code for enzymes that participate in biosynthesis of vitamins and cofactors in *G. m. morsitans*.
2. There is no role of endosymbionts in the expression of *G. m. morsitans* vitamins and cofactors biosynthesis genes.

**1.5 Justification**

Tseste-transmitted African trypanosomes are of major clinical and veterinary importance, hence investigation into biomolecular aspects of tsetse-endosymbiont interaction for development of transmission blockage tools with potential application in control and management of trypanosomiasis is important because of the following reasons. Firstly, chemotherapy is the only other available main line control strategy but is limited by chemoresistance and cytotoxicity, hence there is urgent need for improvement of available and/or development of novel vector control strategies. Secondly, transmission blockage will significantly eliminate the fundamental role of mammalian reservoir hosts including economically important livestock kept by man in epidemiology of the disease, specifically preventing re-infections. Finally, a strategy specific to vector that is environmentally friendly with limited interference with local ecosystem, which the endosymbiont paratransgenesis approach potentially provides, is much needed. In addition, this approach can be applied in other insect borne vector diseases affecting man, livestock and plants.

## 2.1 Trypanosomiasis and African trypanosomes

African trypanosomes (genus *Trypanosoma*) are flagellated parasites that cause sleeping sickness (human African trypanosomiasis, HAT) in man and Nagana (animal African trypanosomiasis, AAT) in cattle. The parasites are transmitted by tsetse flies (genus *Glossina*), the insect vector that facilitates the cyclical transmission of trypanosomes. Tsetse flies are restricted to 37 countries in sub-Saharan Africa, where the disease is also endemic (figure 1).



**Figure 1. Tsetse fly distribution in Africa.** A map of Africa showing the sub-Saharan regions infested by tsetse flies (red), in which African trypanosomiasis is endemic. The black line separates regions where *Trypanosoma brucei gambiense* and *Trypanosoma brucei rhodesiense*, the parasites responsible for Gambian and Rhodesian HAT respectively. The dots represent the approximate cattle distribution, and the whole region has about 70 million inhabitants. This image was adopted from "International atomic energy agency (IAEA) website - http://www.iaea.org/newscenter/features/tsetse/index.shtml

### 2.1.1 Human African trypanosomiasis (HAT)

Sleeping sickness is caused by two subspecies of *Trypanosoma brucei*. *Trypanosoma b. rhodesiense* causes acute form of the disease in East and southern Africa – Rhodesian HAT, while *T. b. gambiense* causes chronic form of the disease in West and Central Africa – Gambian HAT (figure 1) (Holmes, 2013). About 70 million people live in the tsetse belt (figure 1), where there is an estimated 30,000 actual cases of HAT annually (Simarro *et al.*, 2011). New infections reported are on the decline from 9,878 in 2009 to 7,139 in 2010 (Simarro *et al.*, 2011), but this is possibly an underestimation due to geopolitical instability, technical and infrastructural deficiencies in some affected regions and limitation in diagnosis (Odiit *et al.*, 2005; Berrang, 2007). The countries experiencing geopolitical instability include Democratic Republic of Congo (DRC), Angola, Sudan, South Sudan and Uganda (Berrang, 2007).

The two forms of HAT have two clinically distinct stages namely stage 1/early and 2/late. The early stage is when the parasites are localized in the blood and lymph and is characterized by non-specific signs including skin lesions, chancre, pruritus, and cardiac, endocrine, and gastrointestinal problems. Subsequently, the parasites invade the central nervous system (CNS) (stage 2) and is characterized by more serious neurological problems such as tremors, motor weakness, walking difficulties, sensory disorders, visual impairments, headache and sleep disturbances that subsequently deteriorate into coma and eventually death if untreated (Fèvre *et al.*, 2008).

Treatment of both stages is distinct and is associated with cytotoxicity. Suramin is used in treatment of early stage disease; it may cause pyrexia and mild nephrotoxicity, kidney damage, nausea, vomiting, shock, exfoliative dermatitis, severe diarrhoea, and jaundice. Melarsoprol is used to treat late stage HAT because of its ability to cross the blood brain barrier, but it is associated with severe post-treatment reactive encephalitis (PTRE) in 10% of the treated patients of which 50% die due to toxicity (Kennedy, 2006). Side effects of melarsoprol include liver toxicity, severe enterocolitis, fatigue, arthralgia, myalgia and fever, pruritus, urticaria and gastrointestinal reactions, and heart failure. Eflornithine also used for late stage treatment of Gambian HAT may cause convulsions, gastrointestinal reactions, bone marrow toxicity resulting in anaemia, leucopenia, thrombocytopenia and alopecia, fatigue, arthalgia, dizziness, insomnia, fever, headache, and anorexia. These treatment regimens are further limited by drug resistance (Anene *et al.*, 2001) and together with devastating side

effects, some resulting to death, there is need for improvement of current and/ or development of novel control strategies including vector control.

## 2.1.2 Animal African trypanosomiasis (AAT)

Nagana is caused by *T. vivax*, *T. congolense* and *T. brucei* species. *Trypanosoma simiae* is responsible for the disease in pig while *T. evansi* cause surra disease in camels. AAT causes reduced milk and meat production, poor growth rates, reduced calving rates, draught power and manure production (Shaw, 2009). In addition, approximately 50 million cattle are at risk of infection with 3 million deaths occurring annually (FAO, 2013). Since the disease makes the affected areas inhabitable and therefore no economic activities can be undertaken particularly agriculture, the total agricultural loss in terms of Gross Domestic Product is approximately 4.75 billion US dollars per year (FAO, 2013). However, with the disease affecting man hence limiting his other economic activities, in addition to underestimation of infected individuals due to infrastructural and technical limitations (Odiit *et al.*, 2005; Berrang, 2007), the approximated economic impact is most likely a huge underestimate.

## 2.2 Tsetse control

Because of the limitations of chemotherapy (drug resistance and toxicity) and impossibility to develop vaccines due to parasite antigenic variation, vector control has been a preferred alternative (Aksoy, 2003). The adult fly is an accessible target for control since the eggs and the larva are hidden in the female reproductive organs and the pupa development occurs entirely in the soil.

Old methods for controlling tsetse flies were based on elimination of game-reservoir hosts of the parasite, clearing of bush to destroy tsetse breeding sites and creation of fly barriers (Aksoy, 2000) but majority were dropped because of their detrimental effect on the environment (Dransfield *et al.*, 1991). Other methods include use of insecticides, traps and targets, sterile insect technique (SIT) and push-pull method.

## 2.2.1 Insecticides

This involves spraying of residual insecticides like DDT, dieldrin and endosulfan on tsetse infested fields (Wellde *et al.*, 1989) using helicopters. The insecticides kill the tsetse flies over a long period of time. Non-residual insecticides are also used because of their reduced negative effect on the environment as a result of their short half-lives. Aerosolized insecticide sprays mainly target adult flies. The spraying cycles are separated by 16-18 days depending on the temperature (Allsopp and Hursey, 2004) and since the droplets remain suspended in

the air long enough, they kill even emerging flies in subsequent cycles before they start reproducing. Although the method is successful (Kgori *et al.*, 2006), it is not suitable because of a number of reasons, such as it is labour intensive, vector resistance to insecticides, non-target killing of beneficial insects and environmentally unfriendly through pollution causing health risk to spraying workers and the general population (Vreysen *et al.*, 2013).

### 2.2.2 Traps and targets

Both traps and targets attract tsetse flies and eventually kill them. Traps have blue and black coloured cloths that attract tsetse flies. In addition, they have attractants including cattle urine and other chemical attractants such as chloroform. Once on the trap, the flies are directed to a non-return cage where they die of heat or starvation (Brightwell *et al.*, 1991). Targets on the other hand attract the flies and kill them by tarsal contact with insecticides impregnated on the surface of the target (Vale, 1993). These techniques are considered inexpensive and unsophisticated, but are not suitable for use in large areas and for long term (Kappmeier and Nevill, 1999).

### 2.2.3 Push-pull method

Push-pull method uses a chemical repellent (push) on cattle and an attractant (pull) in a trap thus directing the tsetse away from the cattle towards baited traps and targets where they are killed by insecticide, heat or starvation (Hassanali *et al.*, 2008). This method integrates the use of insecticides, chemical repellants and traps which are methods, as already seen, with individual limitations hence making push-pull method also limiting.

### 2.2.4 Sterile insect technique (SIT)

Sterile insect technique is a biological control method that uses irradiated male tsetse flies that are sterile and exploits the fact that females only mate once in their lifetime. Therefore if they mate with a sterile male, the female will not sire offspring thus reducing the tsetse population. This eventually clears their population and reduces transmission of trypanosomes. In 1995, *Glossina austeni* was successfully eliminated from Unguja island of Zanzibar using sterile male tsetse flies and resulted in complete elimination of trypanosomiasis in this geographically isolated area (Vreysen *et al.*, 2000). On the mainland, however, this approach requires area-wide participation in order to be effective, as well as geographically isolated populations to prevent reinvasion of cleared lands (Politi *et al.*, 1995). Importantly, the

success of this approach in Zanzibar increases the hope of successfully using biological control method in other endemic regions.

### 2.2.5 Paratransgenesis

Paratransgenesis is a technique that exploits transformed endosymbiont of tsetse fly for the control of trypanosomiasis. The availability of an *in vitro* culture system for *Sodalis glossinidius* has allowed the use of this commensal symbiont of *Glossina* for the development of a genetic transformation system to introduce and express foreign gene products in tsetse fly (Beard *et al*., 1993). Its localization in the fly tissues such as the mid gut makes it suitable for use to deliver trypanocidal agents and vaccines (Hooper and Gordon, 2001) that directly target the trypanosomes.

*In vitro* transformed *S. glossinidius* are then microinjected into the hemolymph of the tsetse fly and are passed on to F1 generation as well as their offspring while successfully expressing the foreign gene product (Cheng and Aksoy, 1999). These paratransgenic females subsequently maintain infections with recombinant *Sodalis* that continuously express their transgene throughout the life of the fly. Current and future studies include developing methods to increase the stability of transgene expression over time, and identifying novel and effective trypanocidal effector molecules to use in the system (Medlock *et al*., 2013). In addition, improved knowledge on tsetse-endosymbiont-parasite interactions is needed.

### 2.3 Tsetse biology

The tsetse flies belong to the superfamily *Hippoboscoidea* of Diptera which contains four families; *Glossinidae*, *Hippoboscidae*, *Streblidae* and *Nycteribiidae* (Petersen *et al*., 2007). The family *Glossinidae* has only one member (*Glossina*) unlike the other three families in this clade. The genus *Glossina* has 23 species and 8 sub species. These are categorized into three groups named based on the ecological niche they occupy namely *fusca* for forest niche, *morsitans* for savannah niche and *palpalis* for river niche. Common members of *fusca* include *G. fusca fusca*, *G. schwetzi*, *G. brevipalpis* and *G. longipennis*. The *fusca* group inhabit the evergreen forests of west Congo and other than *G. brevipalpis*, other members of this group have no medical or veterinary significance (Leak, 1999). Those of *palpalis* group include *G. p. palpalis*, *G. fuscipes fuscipes* and *G. p. gambiensis*. This group is the vector of *T. b. gambiense* and *T. b. gambiense* that cause HAT. Lastly, those in *morsitans* group that occupy the savannah woodlands include *G. longipalpis*, *G. pallidipes*, *G. m. morsitans*, *G. swynnertoni* and *G. Austeni*, and are the major vectors of *T. b. brucei* which causes Nagana.

### 2.3.1 Life cycle of tsetse fly

Tsetse flies reproduce by adenotropic viviparity (they give birth to a live offspring) and they produce only 8-10 offspring in their 3-4 months lifespan (Attardo *et al.*, 2006a). Female flies usually mate only once in their lifetime then stores the sperms in their spermathecae whereas male flies are promiscuous and can mate many times.

In the female fly, eggs are released periodically from the ovaries and are fertilized. Following fertilization, the egg hatches into a larva which is then retained *in utero* and is nourished on milk secretion from the mother's modified accessory glands until the third instar stage (Attardo *et al.*, 2012). During this feeding process, nutrients along with the tsetse symbionts are transmitted to progeny through the mother's milk gland secretions (Attardo *et al.*, 2006b). The larva is then deposited on soft moist shaded soil approximately 16 days after fertilization (9 days after subsequent fertilizations). The larva never feeds but immediately burrows itself into the soil and pupates within 1-2 hours. This is in contrast to other insects' larvas that feed extensively and store plenty of food before pupating. Therefore, all the nutrients required from the embryo to adult tsetse are derived from the mother (Tobe and Langley, 1978). The adult fly emerges 22-60 days later in a temperature dependent fashion; the warmer the temperatures, the shorter the time it takes for the adult to emerge and the vice versa (Leak, 1999). The offspring female for most *Glossina* species becomes sexually mature 48-72 hour after eclosion (emerging of adult from puparium) while the males take several days (Leak, 1999). Oogenesis seems to be regulated by the presence of a developing larva in the uterus and it takes 6-7 days to complete ( Leak, 1999; Attardo *et al.*, 2006a). Upon ovulation, sperms from the spermathecae fertilize the egg and the cycle begins again. An overview of the life cycle is shown in figure 2.
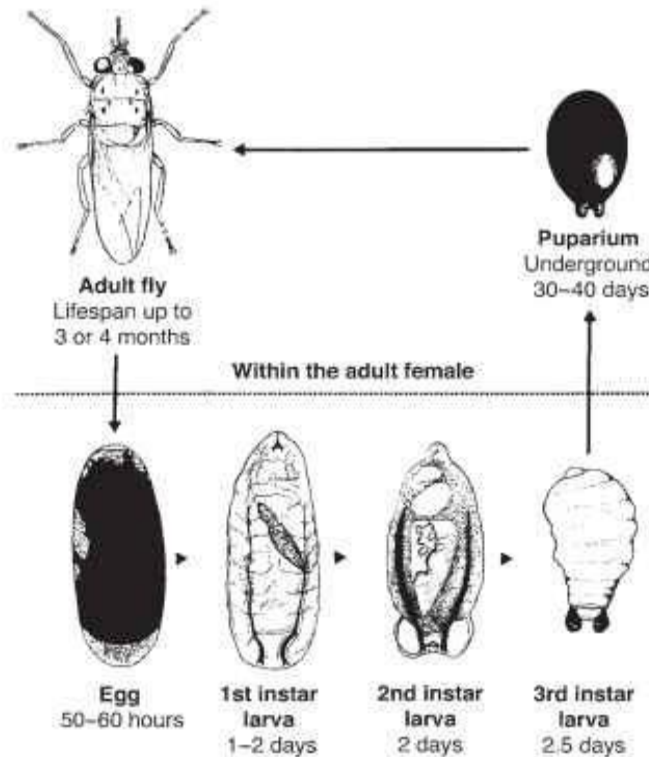
**Figure 2. Life cycle of a tsetse fly.** The egg of mature female is fertilized and develops *in utero* into a 1st followed by a 2nd instar larva. The 3rd instar larva is deposited into the ground and develops into a pupa. Subsequently, the pupae mature into an adult fly and the cycle begins again. This image was obtained from "Tsetse flies (insects), http://what-when-how.com/insects/tsetse-fly-insects/"

### 2.3.2 Tsetse fly feeding

The tsetse has a limited food variety with the young larva depending on the milk secretions only and the adult fly on blood only. The emerged fly usually has some food reserves from its previous stage, but it immediately looks for a blood meal before it can mate (Tobe and Langley, 1978). The exclusive tsetse fly blood diet is deficient in vitamins (Edwards *et al.*, 1957) and this necessitates the fly to supplement its diet possibly from its microbial endosymbionts (Aksoy and Rio, 2005).

The adult tsetse flies are pool feeders and the female can obtain a blood meal weighing several times its own weight. The flies repeatedly penetrate the mammalian host tissue with their proboscis forming a pool of blood on the skin sub-surface from where they suck. Saliva is usually channelled into the wound to prevent the host blood from coagulating on site and this enables the fly to feed longer and it is during this event that an infected fly transmits

(Lehane, 2005) and uninfected fly acquires trypanosomes from infected mammalian host. Unlike the malaria vector, both male and female tsetse flies are obligate hematophages.

Tsetse larva is solely dependent on the nutrients from the mother's milk. This milk is rich in lipids and proteins and is the route of transmission of the bacteria endosymbionts to the larva in the intrauterine stage (Attardo *et al.*, 2012). Since the larva feeds solely on this milk, it is possible that it obtains some of its nutrients from these microbes. Tsetse larva does not feed when it is deposited by the mother like other insects do, but pupates almost immediately implying that all the nutrients it gets from the milk secretions solely originates from the mother's blood meal and possibly the endosymbionts it harbours. Though a blood meal is deficient in nutrients that the tsetse fly requires for its normal physiology, it is possible that the fly supplements this deficiency from the endosymbiont is harbours.

## 2.4 The tsetse genome

Initial studies have shown that the genome size of *Glossina* species vary from 500-600 Mbp (Aksoy *et al.*, 2005a) which is approximately 4 times that of *Drosophila melanogaster* (116.8Mbp) (Misra *et al.*, 2002) and twice that of the malaria vector *Anopheles gambiae* (278Mbp) (Holt *et al.*, 2002). This comparison places *Glossina* at the top and this large genome may provide answers to its unique life cycle and phenotype; the peculiar features include its viviparous reproduction, milk secretion, free-living and obligate hematophagy lifestyle. Notably, tsetse fly is the only insect known to feed its young one on milk, a characteristic of mammalian lifestyle (Attardo *et al.*, 2012). These peculiar characteristics merit *Glossina* a different attention when comparing it to other insects.

## 2.5 Tsetse fly endosymbionts

Insects that rely on one type of meal (such as wood, plant sap or blood) throughout their life often obtain supplement nutrients absent in their core meals by biosynthesis or from microbial symbionts (Douglas, 1989). Symbionts are organisms that a host organism lives with so that the two benefit from each other. Endosymbionts are symbionts that live inside their host and in tsetse flies, the bacterial symbionts implicated are *Wigglesworthia glossinidia*, *Sodalis glossinidius* and *Wolbachia pipientis* (figure 3), and are thought to be the source of vitamins in the fly (Pais *et al.*, 2008).

Intracellular endosymbiont are described as primary (P) because they are required for survival and fecundity of the host. Secondary (S) endosymbionts are facultative bacterial endosymbionts of insects that are often located in syncytial cells near the bacteriocytes and in

various other insect tissue types. They are not essential for host survival and transfer among host individuals and species (Wernegreen, 2002).
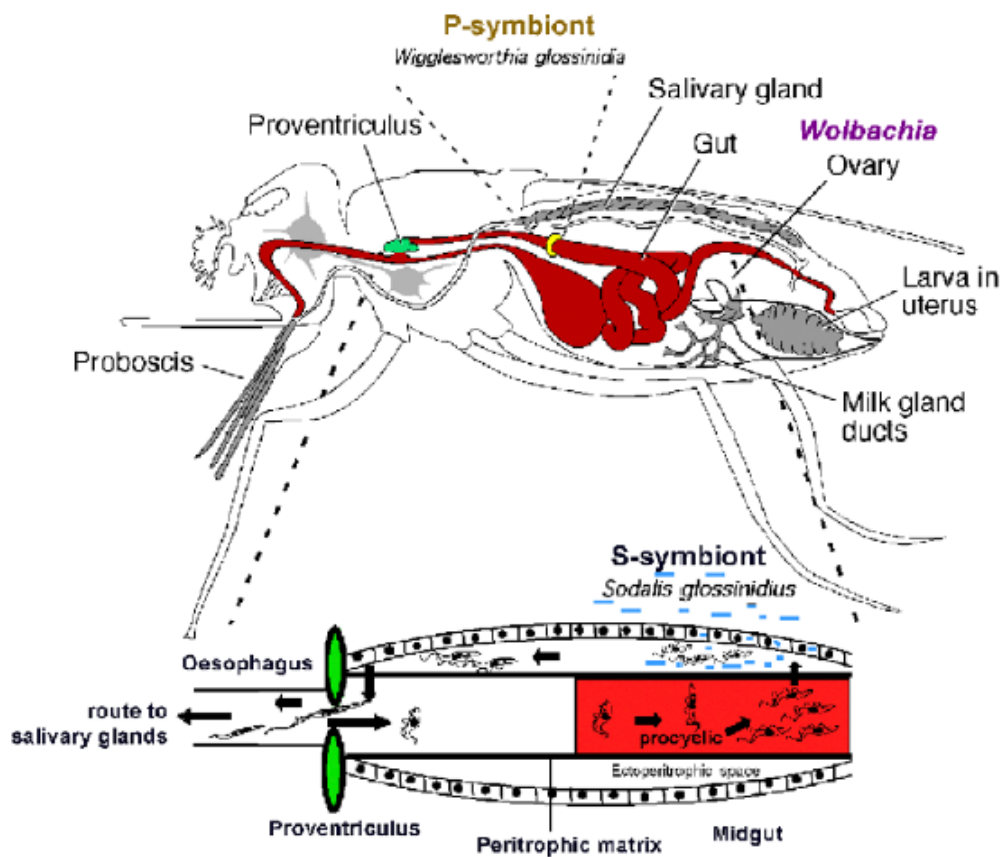


**Figure 3. Association of tsetse fly and endosymbionts.** Intracellular structure of female tsetse fly showing location of various endosymbionts. The primary (P)-symbiont *Wigglesworthia* is found in the bacteriome, an organ in the anterior mid gut. Secondary (S)-symbiont *Sodalis* is found in the gut and hemolymph while *Wolbachia* is in the reproductive tissues. Extracellular forms of *Wigglesworthia* and *Sodalis* are also found in the milk secretions from the milk gland. *Sodalis* associates very closely with trypanosomes in the gut as shown in the enlarged section of the midgut. The image was adopted from Aksoy *et al*. (2005).

*Wigglesworthia glossinidia* and *S. glossinidius* are members of *Enterobacteriaceae* family. *Wigglesworthia glossinidia* is a primary symbiont and it resides inside special epithelial cells called bacteriocytes that make up a U-shaped organ called bacteriome in the anterior gut. It is also found in an extracellular form in the milk secretion (Aksoy and Rio, 2005). *Sodalis glossinidius* is a secondary symbiont and it resides in the mid gut cells (Dale and Maudlin, 1999). The third organism is *Wolbachia pipientis* (O'Neill *et al*., 1993) and it belongs to the

*Rickettsiaceae* family (Doudoumis *et al.*, 2012). Unlike the first two, it is not found in all tsetse species and is found in many other insects like fruit fly, wasps and bees. *Wigglesworthia glossinidia* and *S. glossinidius* are transferred to the offspring via the milk-gland secretions where as *Wolbachia pipientis* is found in the ovaries and is transovarially transmitted through maternal lineages (Aksoy *et al.*, 1997). Generally, the three symbionts are maternally transmitted from one generation to the next.

The role of these symbionts can be studied individually by selectively removing them and determining the resultant effect on the fly. Antibiotics like tetracycline and ampicillin are used for removing the bacteria from the fly. Tetracycline removes all the three symbionts where as ampicillin has effect only on *Wigglesworthia* (Pais *et al.*, 2008). Although the fly does not die, its growth becomes retarded and the rate of fecundity decreases (Nogge, 1976). The ability to reproduce can be partially restored by giving the flies a vitamin B (thiamine, pantothenic acid, pyridoxine, folic acid and biotin) supplemented blood meal (Pais *et al.*, 2008) suggesting that these symbionts play a major role in providing these nutrients.

### 2.5.1 *Wigglesworthia glossinidia*

*Wigglesworthia glossinidia* is an obligate symbiont of tsetse flies that belongs to the gamma-proteobacteria and is believed to have been in association with its host for about 80 million years (Chen and Aksoy, 1999). This endosymbiont resides intracellularly in the bacteriocytes cytoplasm (not surrounded by the host membranes) in the anterior midgut of the tsetse fly. It is located extracellular in the milk gland lumen of tsetse flies (Pais *et al.*, 2008) where it is transmitted between progeny through milk secretions (Attardo *et al.*, 2008). Administration of ampicillin-supplemented blood meal kills extracellular *Wigglesworthia* in the milk glands but does not affect the intracellular forms. However this assures that maternal transmission is not possible therefore the next progeny does not contain *Wigglesworthia* but maintains *Sodalis* (Pais *et al.*, 2008).

The genome of *Wigglesworthia* is 697 kbp in size with a small additional plasmid (pWgb), and has 621 coding sequences (CDS). Ten percent of these CDSs which represents a significant portion of the genome, is involved in the biosynthesis of cofactors, prosthetic groups and carriers (Akman *et al.*, 2002). Specifically, biotin, lipoic acid, riboflavin (FAD), folate, thiamine, pyridoxine, thiazole, pantothenic acid and protoheme metabolism are supported by its genome, supporting the possibility of a role of *Wigglesworthia* in host dietary supplementation (Nogge, 1976).

*Wigglesworthia* is the only tsetse endosymbiont capable of making thiamine based on its genome content *vis-a-vis* that of other endosymbionts (Snyder *et al.*, 2010). It is therefore suggested that it provides thiamine to both the fly and *Sodalis* as genomic evidence suggests that *Sodalis* does not have a thiamine biosynthesis pathway but instead possesses a salvage protein that facilitates acquisition of thiamine (Snyder *et al.*, 2010). This is a good example of metabolic convergence, supporting the hypothesis that microbial symbiont antagonism in the early establishment of dual symbiosis is prevented by functional complementation of symbiont genomes (Moran *et al.*, 2008) as in this case of *Wigglesworthia* and *Sodalis*. Thiamine may play an important role in fecundity of female tsetse since removal of *Wigglesworthia* results in reduced fecundity in the fly (Pais *et al.*, 2008). In addition to dietary supplementation with vitamins to the tsetse, *Wigglesworthia* is also important in the maturation of tsetse immune system (Pais *et al.*, 2008). These suggest that this endosymbiont is essential for the well being of both the fly and other endosymbionts.

### 2.5.2 *Sodalis glossinidius*

*Sodalis* is an enterobacteriaceae found in the mid gut and hemolymph of tsetse flies. In the mid gut it is found in both extracellular and intracellular forms (Aksoy, 2000). Evolutionary relationship studies suggest that its association with *Glossina* is recent unlike with *Wigglesworthia* (Aksoy, 1995; Chen and Aksoy, 1999). This is further supported by the ability to cultivate *Sodalis* on insect cells (Welburn *et al.*, 1987) and cell free media (Beard *et al.*, 1993). This is presumably due to its retained free living lifestyle as it contains most of the genes necessary for free living like other bacteria such as *Escherichia coli K-12, Salmonella,* and *Yersinia* (Toh *et al.*, 2006). In contrast, the other tsetse endosymbionts have not been cultivated outside their host. Another evidence suggesting the recent association is its genome size, which is significantly larger than the genomes of obligate endosymbionts of other insects (Wernegreen, 2002).

The *Sodalis* genome consist of one circular chromosome 4.17 Mbp and an additional three plasmids pSG1, pSG2 and pSG3 as well as a phage ΦSG1. Of the 2,432 putative protein CDSs annotated, an additional 972 are pseudogenes which are homologous to genes involved in defence, transport and metabolism (Toh *et al.*, 2006). This number of pseudogenes is quite high compared to other bacterial species indicating its metabolic adaptation to symbiotic association with its host and the other endosymbionts. All *Sodalis* orthologs are shared with *Wigglesworthia* except for those coding for thiamine, cobalamine, and molybdopterin biosynthesis pathways (Toh *et al.*, 2006). Based on its genome, it is likely that *Sodalis* relies

on its symbiotic association for a number of metabolic process whose ability it has lost through pseudogenization.

### 2.5.3 *Wolbachia pipientis*

Unlike *Wigglesworthia* and *Sodalis*, *Wolbachia* is highly diverse as it infects a wide range of arthropods including at least 65% of insect species (Hilgenboecker *et al.*, 2008). It is also an intracellular and maternally inherited endosymbiont belonging to the α-Proteobacteria. It is found in some tsetse species and therefore not an obligate symbiont of tsetse flies (Werren, 1997). *Wolbachia* infection in its hosts influences reproductive phenotypes in a fashion that enhances its rapid spread. Among these reproductive abnormalities is cytoplasmic incompatibility (CI) that often results in embryonic death because of disruptions in some fertilization events (Hilgenboecker *et al.*, 2008). There is a proposal to use *Wolbachia*-induced CI to suppress agricultural pests and disease vectors (Calvitti *et al.*, 2012). Infected females have a reproductive advantage over their uninfected counterparts as they can produce successful progeny with both the imprinted and normal sperm. This reproductive advantage allows the infected insects to spread into populations and hence transfer a genetic insert of interest competitively against uninfected ones (Sinkins and O'Neill, 2000).

### 2.6 Vitamins and cofactors biosynthesis

Vitamins are the building blocks of cofactors, which are essential in enzyme function. Recently, a variety of vitamins and cofactors biosynthesis mechanisms have been described (Begley *et al.*, 2008). These include thiamine and thiamine pyrophosphate (TPP), riboflavin and flavin adenine dinucleotide (FAD), niacin and nicotinamide adenine dinucleotide (NAD), pantothenate and coenzyme A (CoA), pyridoxine and pyridoxal phosphate (PLP), and folate and tetrahydrofolate (THF) biosynthesis pathways. These biosynthesis pathways in various organisms particularly bacteria and plants have been elucidated. However much remains to be described in insects.

### 2.6.1 Thiamine and thiamine pyrophosphate

Thiamine (vitamin $B_1$) is made up of a pyrimidine and a thiazole molecule joined by a methylene bridge. It is synthesized by most prokaryotes and eukaryotes like yeast and plants. In both prokaryotes and eukaryotes, the thiazole and pyrimidine moieties are made in separate pathways and later joined to form thiamine phosphate as shown in figure 4. A final phosphorylation step gives thiamine pyrophosphate, the active form of the cofactor from thiamine (Jurgenson *et al.*, 2009).
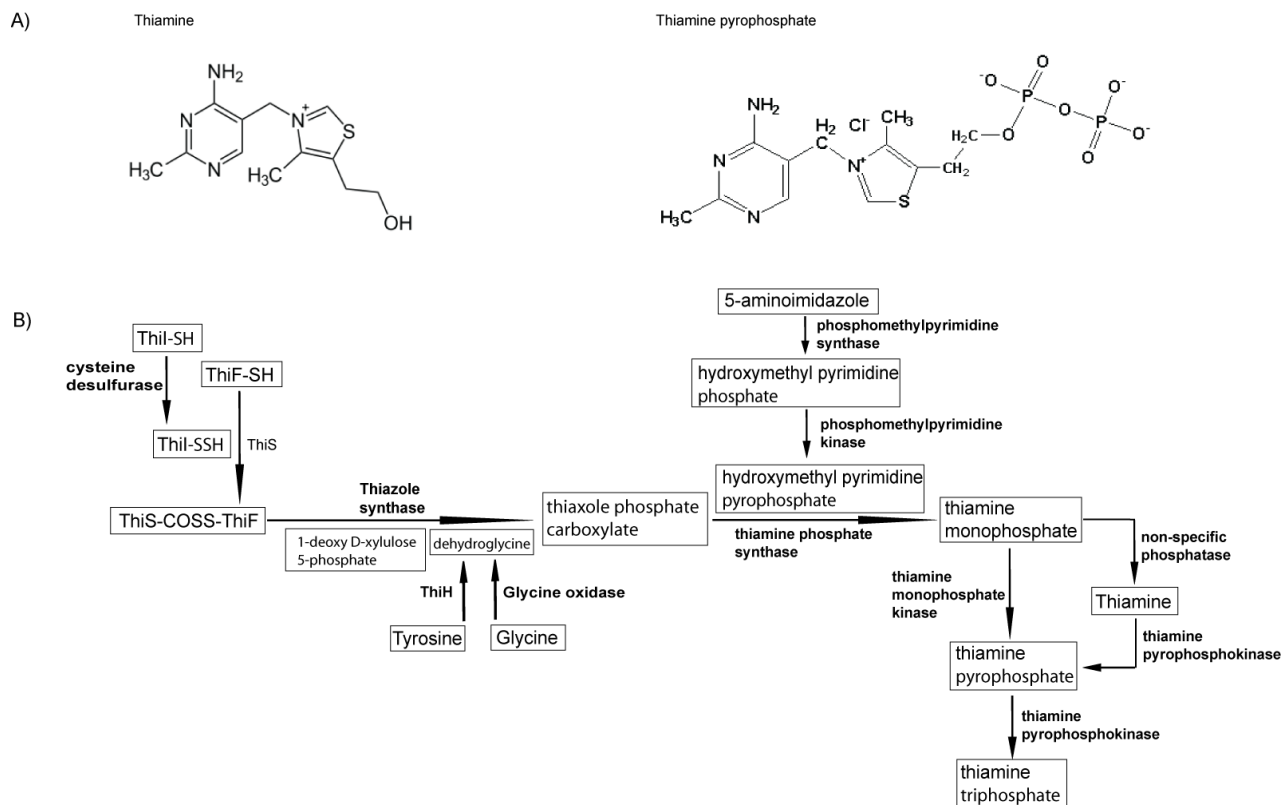
**Figure 4. Thiamine and thiamine pyrophosphate biosynthesis pathway.** Panel **A** shows the structure of thiamine and thiamine pyrophosphate while **B** is the biosynthesis pathway in bacteria. Thiamine's thiazole and pyrimidine rings are joined by a methylene ring. Biosynthesis of thiamine pyrophosphate involves the phosphorylation of thiamine. ThiI-SH,ThiF-SH, ThiH and ThiS represent thiamine biosynthesis proteins I, F, H and S. The figure is adopted from Jurgenson *et al.* (2009).

Bacterial thiamine metabolism is well studied in *E. coli* and in *Bacillus subtilis*; there are limited differences owing to the obligate aerobic nature of *B. subtilis* (Settembre *et al.*, 2004). The formation of the thiazole moiety requires six genes (ThiF, IscS, ThiI, ThiO, TenI and ThiG) where as the pyrimidine requires only two (ThiC and ThiD). The thiazole is made in four distinct steps (Figure 4). First, pyruvate and glyceraldehyde 3-phosphate are coupled together by 1-deoxy-d-xylulose 5-phosphate synthase (Dxs) to form 1-deoxy-dxylulose 5-phosphate (DXP). Second, a sulfur carrier protein called ThiS undergoes adenylylation by ThiF followed by sulfur transfer by ThiI and IscS to yield a thiocarboxy at its C terminus. This sulfur is incorporated into the thiazole ring of thiamine. Thirdly, glycine (*E. coli*) or tyrosine (*B. subtilis*) is converted to dehydroglycine by ThiH or ThiO respectively. Finally, the three products (Thiocarboxy C terminus of ThiS, DXP and dehydroglycine) are coupled together by thiazole synthase (ThiG) to give thiazole phosphate carboxylate tautomer. The

enzyme TenI in *B. subtilis* then aromatizes the tautomer to thiazole phosphate carboxylate (Begley *et al.*, 2008). The pyrimidine moiety of thiamine is formed by a rearrangement reaction of aminoimidazole ribotide catalysed by the gene product of the ThiC and ThiD gene (Lawhorn *et al.*, 2004).

## 2.6.2 Riboflavin, flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN)

Riboflavin is the building block of two cofactors, flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN). These cofactors are used in some dehydrogenation reactions e.g. FAD in the tricarboxylic acid (TCA) cycle for ATP production.
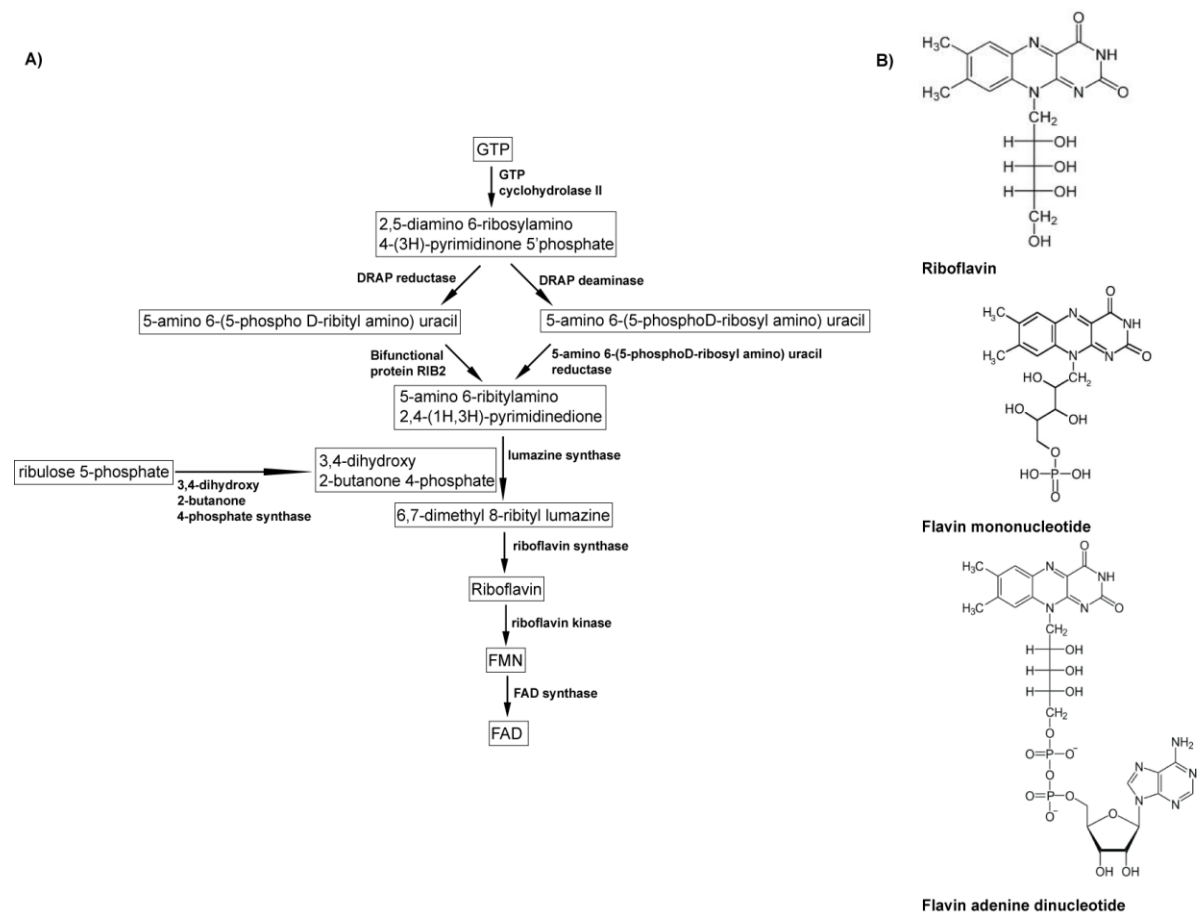


**Figure 5. Riboflavin, flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) biosynthesis.** The pathway (**A**) begins from ATP and involves the synthesis of the lumazine molecule by lumazine synthase. The lumazine molecule is then converted to riboflavin from which FMN and FAD are generated. **B** shows chemical structures of riboflavin, FMN and FAD molecules. Abbreviations: GTP, guanosine triphosphate; DRAP, 2,5-diamino 6-ribosylamino 4-(3H)-pyrimidinone 5'phosphate. Pathway was adopted from Marx *et al.* (2008).

As shown in figure 5, its synthesis begins with opening of the guanosine triphosphate (GTP) imidazole ring by GTP cyclohydrolase II too yield 2,5-diamino-6ribosylamino-4-(3H)-pyrimidinone-5'-phosphate (Schramek *et al.*, 2001). This product is subsequently converted to 5-amino-6-ribitylamino-2,4-(1H,3H)-pyrimidinedione, which is condensed with 3,4-dihydroxy-2-butanone-4-phosphate by the enzyme 6,7-dimethyl-8-ribityllumazine synthase to form 6,7-dimethyl-8-ribityllumazine. Riboflavin is then formed from 6, 7-dimethyl-8-ribityl lumazine by riboflavin synthase (Fischer *et al.*, 2002). FMN and FAD are then made by addition of nucleotides to riboflavin.

### 2.6.3 Nicotinamide adenine dinucleotide (NAD)

Nicotinamide adenine dinucleotide ($NAD^+$) is a cofactor required in all organisms most notably in glycolysis and TCA cycle where it is interchanged between $NAD^+$ and its reduced form (NADH) to generate ATP. It is also a substrate of some enzymes such as ADP-ribosyl transferase and ADP-ribosyl cyclases.

It is biosynthesized through two main pathways – first, *de novo* biosynthesis pathway by plants and some bacteria and second, the salvage pathway by animals and some plants, fungi and bacteria (Lin *et al.*, 2010). The *de novo* pathway involves five enzymes and it begins from the amino acid L-aspartate as shown in figure 6 below. This is converted to quinolinate in two steps by two enzymes, L-aspartate oxidase and quinolinate synthetase. Quinolinate is then converted to nicotinate mononucleotide and subsequently to nicotinate adenine dinucleotide by quinolate phosphoribosyltransferase and nicotinate/ nicotinamide mononucleotide adenylyltransferase respectively. Lastly nicotinate adenine dinucleotide is converted to $NAD^+$ by $NAD^+$ synthetase.

Salvage pathway can either be a two-step reaction that begins from nicotinamide as in mammals and some bacteria or a four-step reaction that also begins from nicotinate as in some bacteria, plants and fungi. In the two-step reaction, nicotinamide is converted to nicotinamide mononucleotide by nicotinamide phosphoribosyltransferase and eventually converted to $NAD^+$ by nicotinate/ nicotinamide mononucleotide adenylyltransferase. In the four step reaction, nicotinamide is converted to nicotinate by nicotinamidase. Nicotinate is then converted to nicotinate mononucleotide which joins the *de novo* pathway to proceed into a two-step reaction that generates $NAD^+$.
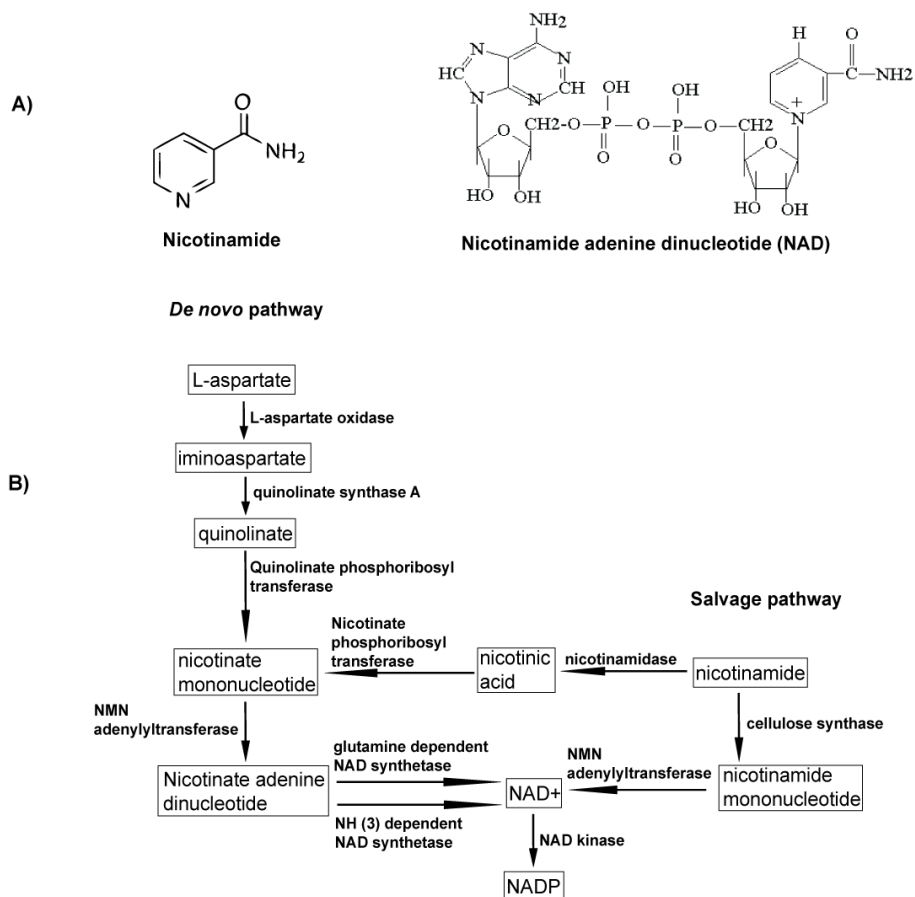
**Figure 6.** *De novo* **and salvage biosynthesis of nicotinamide adenine dinucleotide (NAD).** Panel A shows the chemical structure of NAD. In the pathway (B) biosynthesis in plants and bacteria begins from L-aspartate while the salvage pathway for some plants, bacteria and fungi begins from nicotinic acid. In animals, the salvage pathway begins from nicotinamide. This diagram is adopted from Lin *et al.* (2010).

### 2.6.4 Pantothenate and coenzyme A

Coenzyme A is synthesized from pantothenate (vitamin B5) and it functions as a cofactor in energy metabolism and metabolism of lipids where it acts as a carrier of acyl groups. In bacteria, its biosynthesis begins with synthesis of β-alanine from L-aspartate. On another branch of the pathway, α-ketoisovalerate is synthesized from L-valine then undergoes hydroxymethylation followed by decarboxylation to form pantoate. Pantothenate is then formed by a branched-chain-amino-acid transaminase from pantoate and β-alanine. Coenzyme A is then formed from the vitamin in the following steps. Phosphorylation followed by addition of cysteine to pantothenate and decarboxylation forms 4-phosphopantothenine. This is subsequently adenylated to form dephospho coenzyme A and

20

finally phosphorylation by dephopho coenzyme A kinase forms coenzyme A (Begley *et al.*, 2001). These steps are elaborated in the figure 7 below.
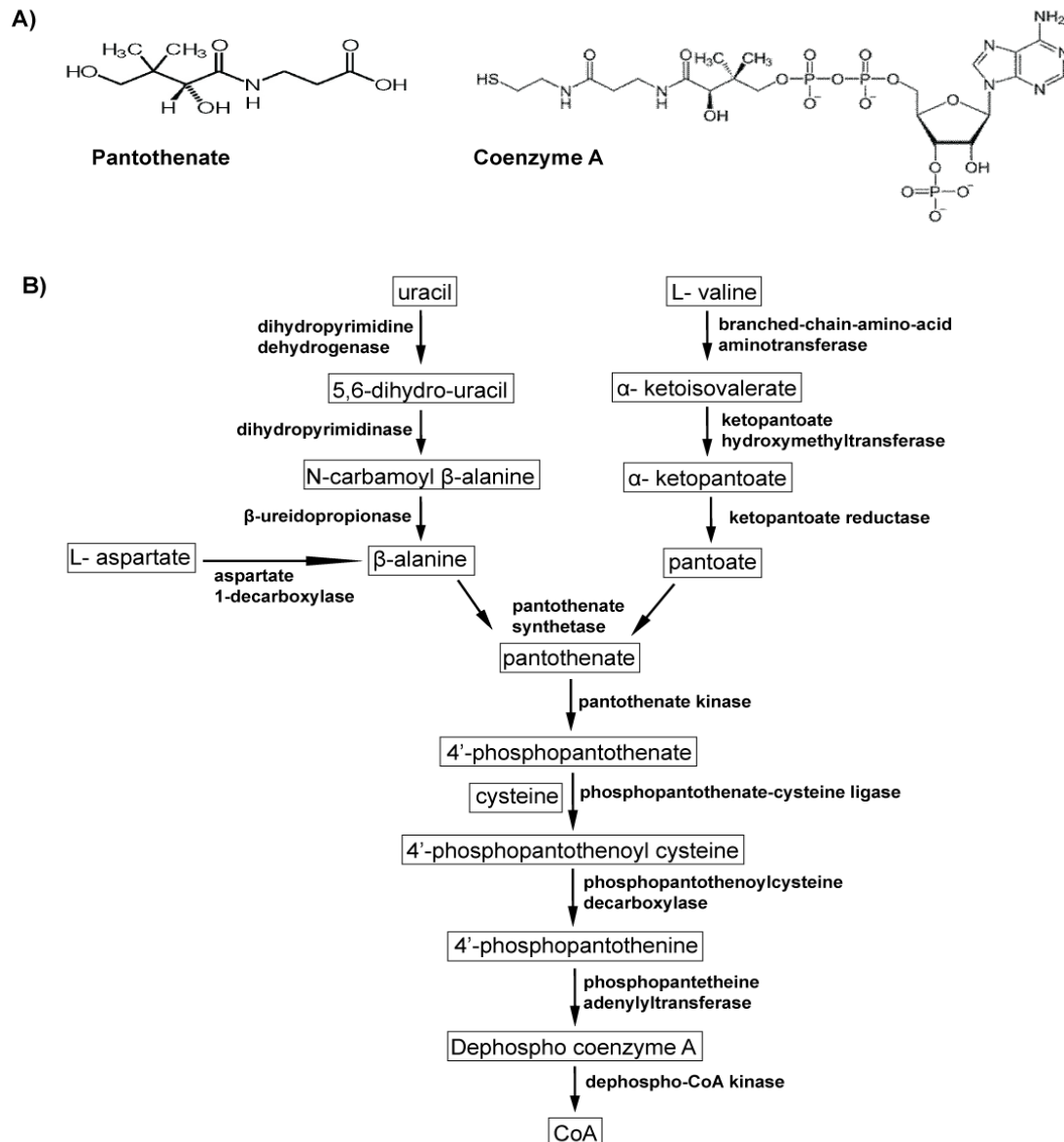


**Figure 7. Coenzyme A biosynthesis in bacteria.** Panel **A** shows the chemical structure of pantothenate and coenzyme A. In the pathway (**B**), bacteria derive *β*-alanine from L-aspartate while animals derive it from uracil. Pantothenate is then made by adding *β*-alanine to pantoate, and is subsequently used to generate coenzyme A. This pathway is adopted from Begley *et al.* (2001).

## 2.5.5 Pyridoxine and pyridoxal phosphate

Pyridoxine also known as vitamin B6 is the precursor for pyridoxal phosphate (PLP), a cofactor involved in transfer of amino group in a variety of enzymes involved in amino acid metabolism. Synthesis of this cofactor has two pathways, a *de novo* and a salvage pathway.

21

The precursor metabolites for the *de novo* pathway are D-erythrose-4-phosphate and the glycolysis intermediate glyceraldehyde-3-phosphate. D-erythrose-4-phosphate undergoes four enzymatic steps as shown in the figure 8, to form pyridoxine phosphate that is eventually condensed with glyceraldehyde-3-phosphate to form PLP. One of the salvage pathways involves the conversion of pyridoxine directly to PLP by pyridoxine kinase (2.7.1.35). Another involves conversion of pyridoxamine by pyridoxine kinase to form pyridoxamine monophosphate which is eventually phosphorylated to form PLP.
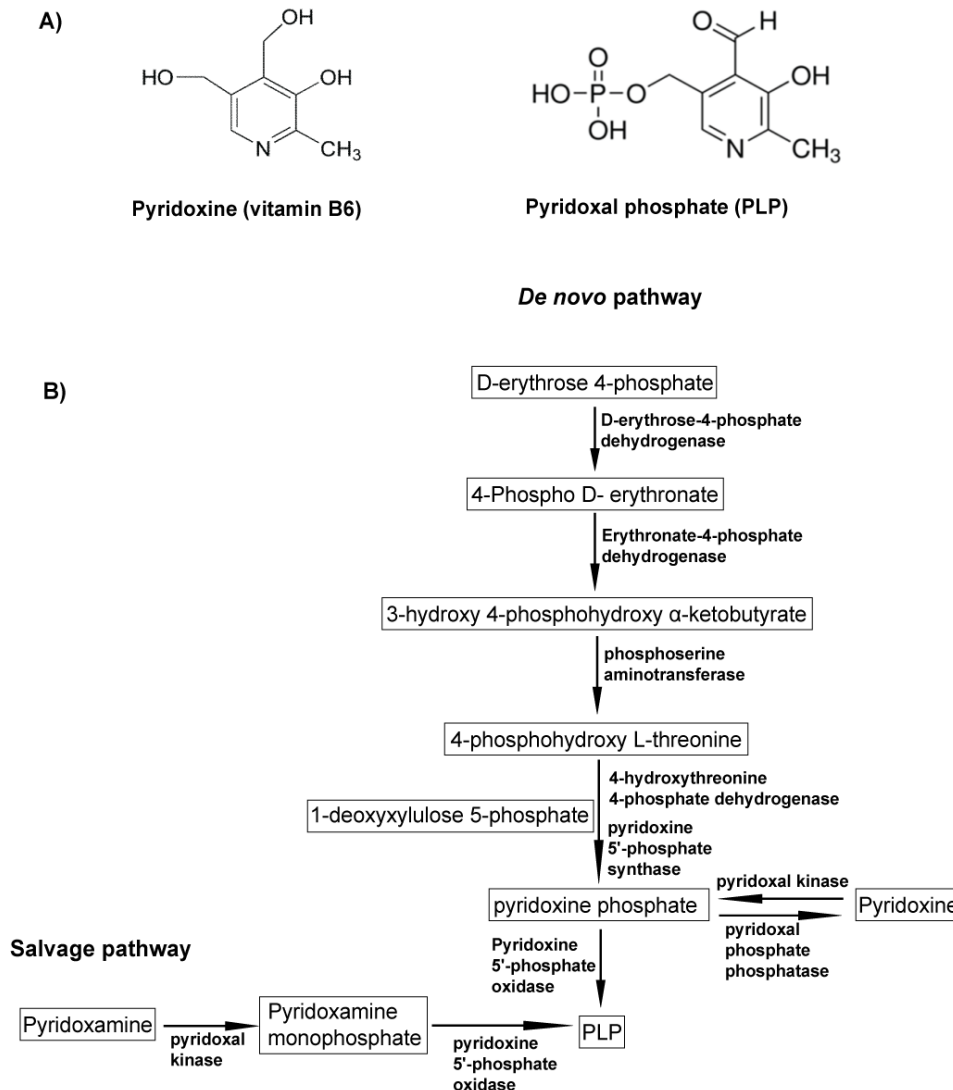


**Figure 8. *De novo* biosynthesis and salvage of pyridoxine and pyridoxal phosphate (PLP).** In *de novo* biosynthesis bacteria synthesize pyridoxine from D-erythrose 4-phosphate and 1-deoxyxylulose 5-phosphate. PLP is then synthesized by phosphorylation of the ketone form of pyridoxine. In the salvage pathways in animals, PLP is synthesized by phosphorylation of pyridoxamine or pyridoxine. The figure is adopted from Tanaka *et al.* (2005).

### 2.5.6 Biotin

Biotin is also known as vitamin B7 and is a key cofactor that carries carbon dioxide in carboxylation, decarboxylation and transcarboxylation reactions. Cellular processes such as amino acid metabolism, fatty acid synthesis and gluconeogenesis require biotin as a carrier of carboxyl group. The biosynthesis pathway can be divided into two stages with the first stage involving the synthesis of a pimelate moiety, which is used in the second stage to make the ring portion of biotin.

The first stage utilizes fatty acid synthetic enzymes to grow a malonate moiety into a pimelate. The second stage begins by conversion of pimelate thioester to 8-amino-7-oxononanoate using 8-amino-7-oxononanoate synthase. Then this undergoes transamination to form 7,8-diaminononanoate. Dethiobiotin is subsequently formed by dethiobiotin synthase and finally biotin is formed from dethiobiotin though the action of biotin synthase. The biotin is eventually ligated to a lysine residue of a biotin dependent enzyme by biotin protein ligase.
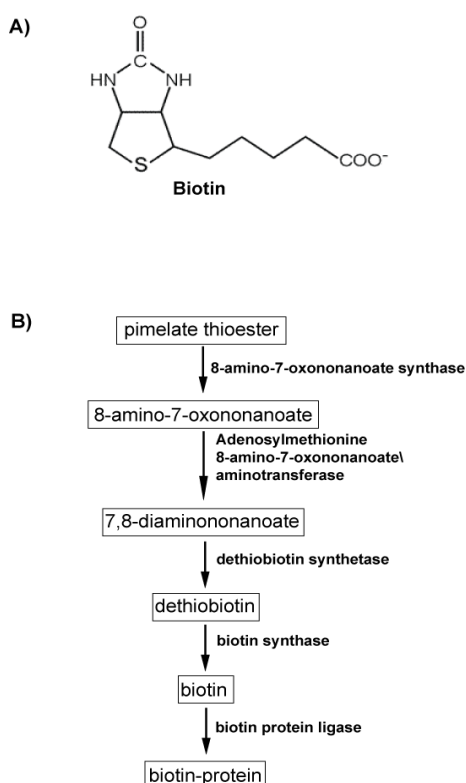
**Figure 9. Biotin biosynthesis. A** shows the structure of biotin. **B.** Biotin biosynthesis involves the rearrangement of pimelate thioester from the fatty acid metabolism in a four step reaction. Biotin is then ligated to an enzyme where it acts as a cofactor. The pathway was adopted from Lin and Cronan, (2011).

### 2.5.7 Folate and tetrahydrofolate

Folate (vitamin B9) is important for making the cofactor tetrahydrofolate (THF). THF is used in enzymatic reaction as a one-carbon carrier in enzymes such as those involved in amino acid, nucleic acids and pantothenate synthesis (Cossins, 2000). It is synthesized from pterin, para-aminobenzoic acid (PABA) and glutamate as shown in figure 10.

Pterin synthesis begins with the conversion of GTP to dihydroneopterin triphosphate by GTP cyclohydrolase I, followed by two dephosphorylation steps, - the first carried out by dihydroneopterin triphosphate pyrophosphatase (Gabelli *et al.*, 2007) and the second involves a non-specific phosphatase as in *E. coli* (Suzuki and Brown, 1974). However, according to KEGG pathways (Kanehisa *et al.*, 2012), alkaline phosphatase is capable of dephosphorylating dihydroneopterin triphosphate directly to dihydroneopterin. The resulting dihydroneopterin is then broken down into glycolaldehyde and 6-hydroxymethyldihydropterin by dihydroneopterin aldolase. The later is then phosphorylated into 6-hydroxymethyldihydropterin pyrophosphate by 6-hydroxymethyldihydropterin pyrophosphokinase.

PABA is synthesized from chorismate in two steps. Aminodeoxychorismate synthase converts chorismate to aminodeoxychorismate which is then converted to PABA by aminodeoxychorismate lyase. PABA and the pterin are conjugated by dihydropteroate synthase (EC: 2.5.1.15) to form dihydropteroate which forms dihydrofolate on addition of glutamate as shown in the figure 10.
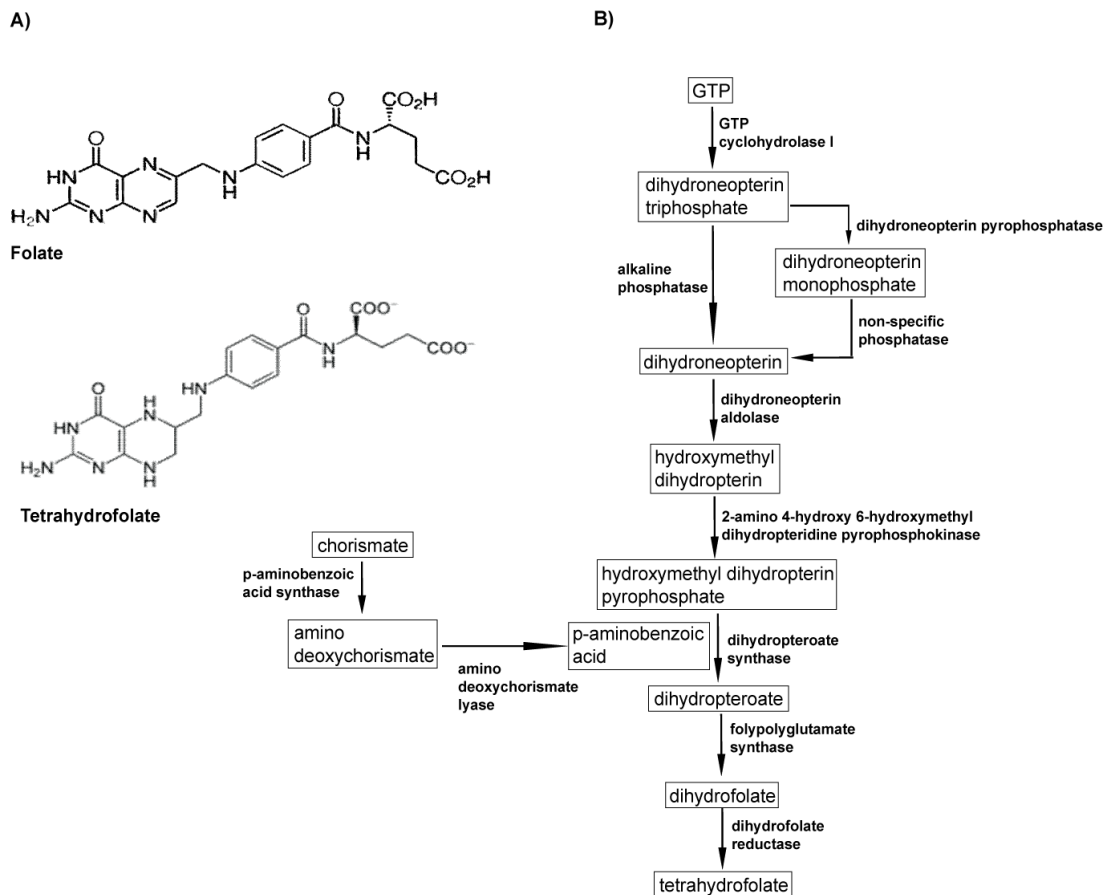
**Figure 10. Folate and tetrahydrofolate biosynthesis. A** shows the structure of folate and tetrahydrofolate (THF). **B.** Animals, plants and bacteria synthesize folate from para-aminobenzoic acid (PABA), glutamine and a pterin ring. Tetrahydrofolate (THF) is then generated from reduction of dihydrofolate. This pathway was adopted from Hanson and Gregory, (2011).

THF is lastly formed in a reduction reaction mediated by dihydrofolate reductase. Glutamates can be added to the THF by formylpolyglutamate synthase to form a polyglutamate tail and this is the preferred cofactor form while the monoglutamate form is preferred by transport proteins (Hanson and Gregory, 2011).

## 2.7 Vitamins and cofactors biosynthesis in tsetse symbionts

The small genome of *Wigglesworthia* has retained the ability to synthesize various vitamin metabolites including biotin, thiazole, lipoic acid, riboflavin, folate, pantothenate, thiamine, pyridoxine, protoheme, and nicotinamide (figure 11) (Akman *et al.*, 2002) which are known to be low in the single diet of tsetse - vertebrate blood. *Wigglesworthia* is thought to be the main source of thiamine in the tsetse-endosymbiont association given that *Sodalis* salvages (Snyder *et al.*, 2010).
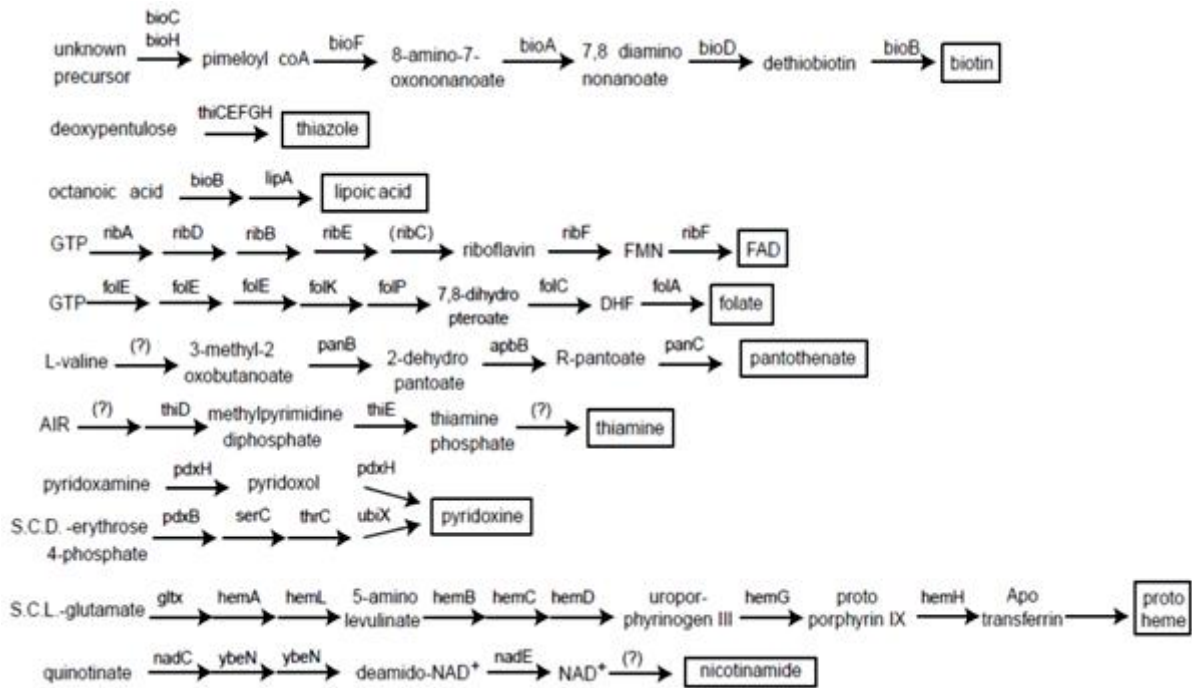
**Figure 11. An overview of cofactors biosynthesis in *Wigglesworthia*.** B-vitamins and cofactors synthesized by *Wigglesworthia* are biotin, thiamine, riboflavin, folate, pantothenate, pyridoxine and nicotinamide. However *Wigglesworthia* lacks some of the genes that encode enzymes (marked with "(?)") in biosynthesis pathways for pantothenate, thiamine and nicotinamide. This image was adopted from Akman *et al.* (2002).

In addition to thiamine, *Sodalis* also lacks biosynthetic pathways for cobalamine, and molybdopterin biosynthesis pathways which are present in *Wigglesworthia* (Toh *et al.*, 2006). Sodalis however appears to have the potential to biosynthesize folate, riboflavin, pantothenate, biotin and nicotinamide. The initial reactions of pyridoxine biosynthesis pathway appear incomplete (figure 12).

*Wolbachia* is not an obligate mutualist in tsetse flies and is not found in all populations of flies (Werren, 1997). It is actually considered as a parasite more than a symbiont. Its genome shows a loss of most vitamin biosynthesis pathways and therefore its role in provision of these nutrients may not be significant.
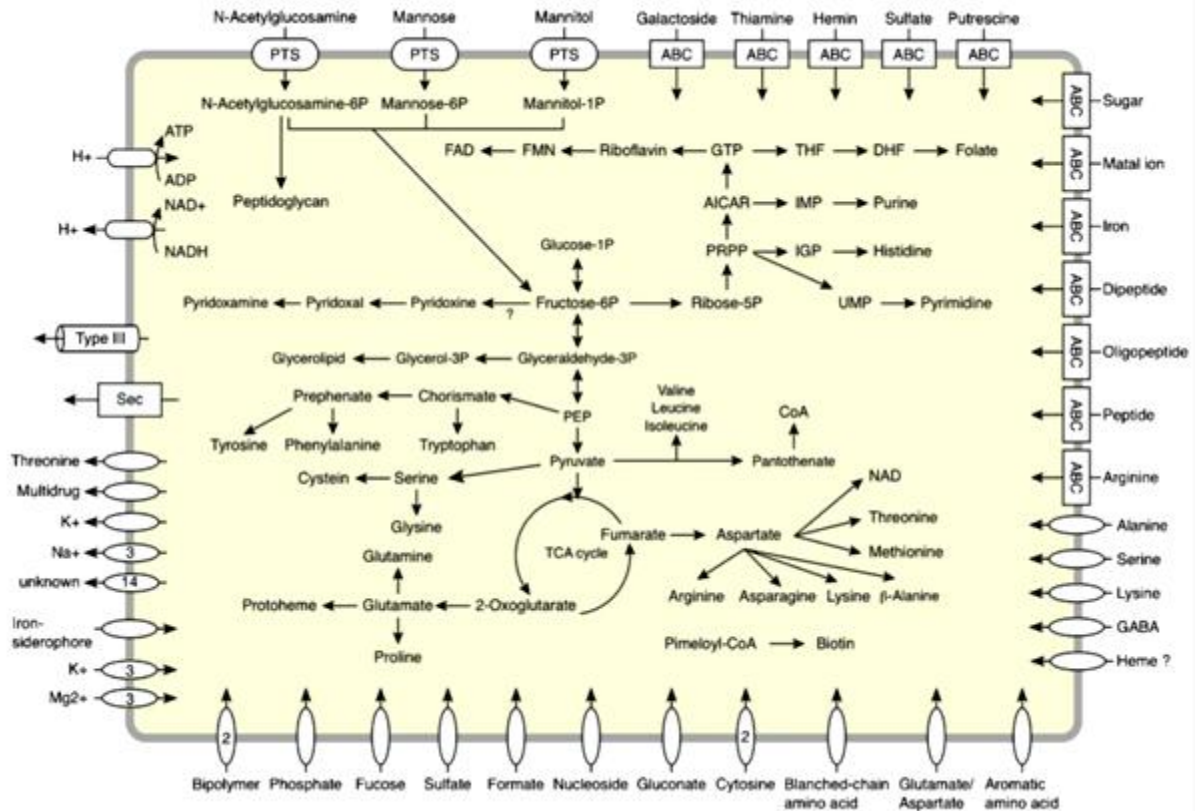
**Figure 12. An overview of *Sodalis glossinidia* metabolic pathways including B-vitamins and cofactors biosynthesis pathways.** *Sodalis glossinidia* does not synthesize thiamine and instead salvages using an ABC transporter. Its pyridoxine biosynthesis pathway also appears to be incomplete. The image was adopted from Toh *et al.* (2006).

**Summary**

African trypanosomes transmitted by tsetse flies cause the deadly disease called African trypanosomiasis in both man and his animals leading to a huge economic loss. Control of trypanosomiasis is mainly by chemotherapy and vector control. The drugs used in human treatment include pentamidine, suramine, eflornithine and melarsoprol. Most of these drugs are old, dating back to 1950 and are limited by chemoresistance and drug toxicity. In addition, diagnosis suffers various limitations including low sensitivity and specificity, inapplicability in field environment and high cost. (Odiit *et al.*, 2005; Kennedy, 2006; Berrang, 2007; Fèvre *et al.*, 2008).

Vector control involves use of insecticides, traps and targets and sterile insect technique but these methods have their individual limitation therefore necessitating improvement and/ or development of novel strategies (Vale, 1993; Politi *et al.*, 1995; Kappmeier and Nevill, 1999; Kgori *et al.*, 2006). One potential strategy is paratransgenesis that exploits the endosymbionts

in tsetse fly to either control tsetse populations or prevent the fly from transmitting trypanosomes (Beard *et al.*, 1993). In this strategy, transgenic endosymbionts can be propagated in trypanosomes infected tsetse flies where they express a trypanocidal agent that kills trypanosomes and thus preventing transmission. This method is under development but the process is hindered mainly by the limited knowledge available on the tsetse-endosymbiont interaction (Medlock *et al.*, 2013). The method is favourable since it only targets the trypanosomes and does not affect the fly and is thus environmentally friendly.

One major interaction between tsetse and its endosymbionts is in provision of nutrients deficient in tsetse blood meal i.e. vitamins and cofactors (Snyder *et al.*, 2010). This is however speculative since some biosynthesis pathways for vitamins and cofactors in endosymbionts are incomplete because of absence of some enzymes due to loss of their respective genes i.e. genome reduction (Akman *et al.*, 2002; Toh *et al.*, 2006; Nikoh *et al.*, 2011). In addition, the biosynthesis pathways in tsetse fly are unknown and therefore this study sought to determine the fly's biosynthesis pathways and the nature of interaction with endosymbionts in biosynthesis of vitamins and cofactors.

As revealed by bioinformatics analysis, the genome of *G. m morsitans* has enzymes that are missing in the endosymbionts B-vitamins biosynthesis pathways and further encodes full cofactor biosynthesis machinery. A select number of genes in the pathways examined were confirmed to be authentic by polymerase chain reaction (PCR) amplification and sequencing. In addition, potential interaction at molecular and biochemical level is suspected due to differential gene expression in presence and absence of endosymbionts, specifically *Wigglesworthia*. Upregulation of expression in absence of endosymbiont indicates compensation for deficit in some metabolites shared with endosymbiont, implying interaction.

# CHAPTER THREE

# MATERIALS AND METHODS

## 3.1 The genome sequence of *Glossina morsitans morsitans*

The genome sequence of *G. m. morsitans* Yale strain 1 used in this study was sequenced and assembled at the Wellcome Trust Sanger Institute, UK (IGGI, 2014). DNA was extracted from female flies and their female progenies. The mother fly was pre-treated with tetracycline to remove all bacteria symbionts following the method developed by Pais *et al.* (2008). This genome assembly consisted of 366 Mega base pairs (Mbps) with 12,308 predicted proteins and is available for public access at VectorBase (http://www.vectorbase.org/). In this repository, the genome is referred to as Glossina-morsitans-Yale_SCAFFOLDS_GmorY1.fa and it contains 13,807 sequences and a total of 366,195,856 letters.

## 3.2 Transcriptome data for annotation and gene expression

Three transcriptome data used in this study were a kind gift from Prof. Serap Aksoy of Yale University, USA; total RNA was used to prepare complementary DNA (cDNA) which represented the total expressed genes i.e. mature RNA. The first data consisted of a mixture of reads generated using RNA from the midgut, salivary glands, fat body and reproductive tissues and this was used for manual annotation of genes. The second and third data were obtained using RNA from the midgut of ampicillin treated and untreated adult female flies respectively. These were used to determine differential gene expression.

## 3.3 Identification of global ortholog genes

For accurate identification of tsetse B-vitamin and cofactor biosynthesis enzymes, experimentally validated orthologs were first identified in eukaryotes and prokaryotes. This was carried out through extensive literature searches at PubMed (http://www.ncbi.nlm.nih.gov/pmc/) to identify experimentally validated enzymes and interrogation of organism specific databases. The targeted organisms and their respective databases include *Escherichia coli* (http://www.ecogene.org/), *Bacillus subtilis* (http://bsubcyc.org/), *Drosophila melanogaster* (http://flybase.org/), *Arabidopsis thaliana* (http://www.arabidopsis.org/), *Saccharomyces cerevisiae* (http://www.yeastgenome.org/), and *Homo sapiens* (http://www.genenames.org/). Subsequently, the panel of experimentally validated orthologs were analysed for domain organization at public protein databases namely pfam (Punta *et al.*, 2012), prosite (Sigrist *et al.*, 2013) and interpro (Hunter *et al.*,

2012), and polypeptide physicochemical properties i.e. protein lengths, isoelectric points (pI) and molecular weight (MW) at expasy http://web.expasy.org/compute_pi/ (Wilkins *et al*., 1999). After robust analysis, the identified proteins were used to identify orthologs in other organisms namely *Anopheles gambiae* (http://vectorbase.org/), *Tribolium castenum* (http://beetlebase.org/), *Wigglesworthia glossinidia*, *Sodalis glossinidius*, *Wolbachia pipientis* (http://www.ncbi.nlm.nih.gov/genbank/) and *Sorghum bicolour* (http://www.plantgdb.org/SbGDB/). The proteins identified were analysed as above. Together, this provided a global outlook of all the possible enzymes, their physicochemical properties and domain architectures, and pathways involved in vitamin B family and cofactor biosynthesis. These enzymes were used to search the *Glossina* genome.

## 3.4 Homology search for *Glossina* B-vitamins and cofactors biosynthesis genes

To identify the respective orthologs in *G. m. morsitans*, BLAST searches were carried out using the identified experimentally and computationally validated enzymes as query sequences at VectorBase. To ensure robust analysis and prevent bias, a criterion was set for subsequent analysis. First, all identified enzymes from various organisms were used to search the *Glossina* genome. Second, all the low e-value hits were interrogated before rejection as orthologs. Third, all the hits were analysed for protein domain organization and physicochemical properties as described earlier. Fourth, *Glossina* proteins identified multiple times, and have conserved domain architecture were considered true orthologs. Fifth, the identified *Glossina* proteins were used in reverse search to determine if they identify the initial query proteins. Finally, to prevent reporting an endosymbiont ortholog as tsetse genes, the *G. m. morsitans* yale strain 1 genome was from tsetse treated with tetracycline to remove all endosymbionts. In addition, vitamins and cofactor biosynthesis for endosymbionts were also determined hence known. These ensured robust analysis to identify tsetse orthologs involved in vitamins and cofactors biosynthesis.

## 3.5 Comparative analysis

Tsetse fly is a hematophagous parasite-transmitting insect. For comparative analysis, orthologs from a similar hematophagous parasite-transmitting insect, namely *Plasmodium* transmitting *Anopheles gambiae* (feeds on blood and plant sap) and non-hematophagous feeders namely *D. melanogaster* (ripe fruits) and *T. castanum* (flour) were used. In addition, since endosymbionts have been suggested to be involved in vitamin metabolism (Pais *et al*., 2008), bacterial endosymbionts namely *Wigglesworthia glossinidia*, *Sodalis glossinidius*, *Wolbachia pipientis*, as well as free-living bacteria *E. coli* and *B. subtilis* were included.

These enabled integration of possible molecular and biochemical interactions between tsetse fly and its endosymbionts, and a comparative analysis with other systems.

## 3.6 Gene modelling

The results obtained from the homology search were structurally annotated using Artemis genome browser (Carver *et al.*, 2008) and CLC genomics workbench 4.8. Artemis was used to determine the genes start (ATG) and stop (TAA, TAG and TGA) codons and also correct splice boundaries (GT-AG) as illustrated in figure 13. Further, to assure that the gene coding regions (exons) have the same sequence as their respective mRNA and the computationally identified genes are accurate, validation by mapping illumina RNA sequences (cDNA) using CLC genomics workbench was carried out.
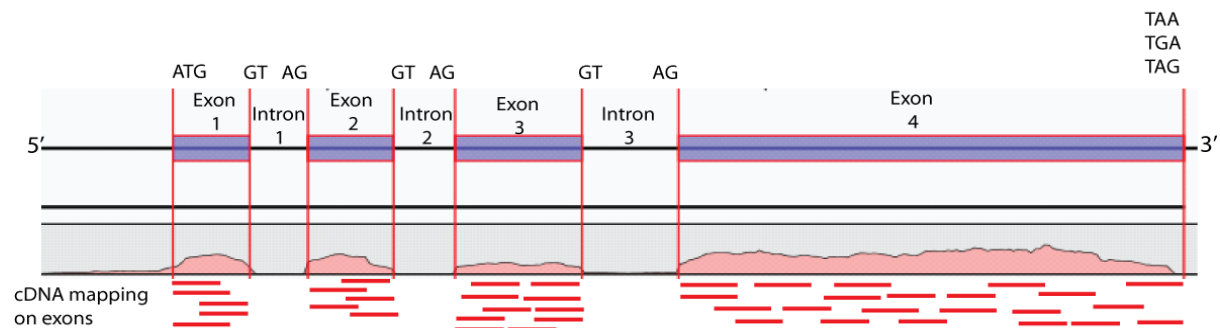


**Figure 13. Complementary DNA (cDNA) mapping on a gene model.** From the 5' end, the coding sequence begins with a start codon, ATG, and ends with a stop codon TAA, TGA or TAG at the 3' end. The exons shown in purple fill are separated by introns (uncoloured black lines between exons). The exon-intron boundaries have consensus sequences GT-AG. The RNAs - shown as short red lines - from RNAseq data do not map on introns since they are generated from mature mRNA.

## 3.7 Assignment of enzyme commission numbers

The *Glossina* protein sequences were assigned enzyme commission (EC) numbers based on homology (Lee *et al.*, 2007) and domain similarity to experimentally validated protein sequences from swissprot as follows. First, EC numbers were assigned to sequences based on identity of ≥60% (Tian and Skolnick, 2003) to their best blastp hit in the nr database and subsequently annotated using blast2GO (B2G) annotation tool (Conesa *et al.*, 2005) at an e-value of 1.0e-6. This tool assigns EC numbers by blasting protein sequences against the gene ontology (GO) database. Since the database used was made up of swissprot proteins, the

sequences that were assigned EC numbers were orthologous to experimentally validated and manually curated protein sequences.

Secondly, the domain potentially representing the active site of the enzyme had to be present to validly assign a specific EC number. Orthologous sequences that lacked any domain were regarded as pseudo genes while those that could not be assigned EC numbers using B2G but had the required domain were assigned putative function using phylogeny (Pellegrini *et al.*, 1999). Phylogenetic analyses were done using phylomeDB tool available at http://phylomedb.org/. This is a collection of maximum likelihood phylograms created using genes from *Glossina*, other dipterans and outgroups. The sequence was assigned the EC number of the proteins that cluster with it in the phylograms.

### 3.8 Expression analyses

To determine the expression of particular genes, RNA was mapped on these gene models using CLC genomics workbench (see figure 13). The two transcriptome datasets obtained from the midgut of ampicillin treated flies were analyzed using the RNA-seq (Mortazavi *et al.*, 2008) analysis tool in CLC genomics workbench. The analysis parameters were made more stringent by not allowing non specific mapping of reads to genes. Subsequently, differential gene expression was determined by comparing expression (RNA levels of selected genes) from ampicillin treated and untreated flies. The expression of vitamin transport proteins in *Glossina* was also determined in a similar way. Gene expression values were calculated in terms of reads per kilobase pair per million reads (RPKM).

$$RPKM = \frac{Total\ exon\ reads}{mapped\ reads\ (millions)\ X\ \ exon\ length\ (Kb)}$$

Total exon reads is the number of cDNA fragment mapped on a gene that fall entirely within an exon or in exon-exon or exon-intron junctions. Mapped reads is the total number of all the reads that have mapped to all the individual genes.

### 3.9 Gene validation

To demonstrate that the genes identified *in silico* using bioinformatics tools actually exist, polymerase chain reaction (PCR) amplification of selected genes from *G. m. morsitans* whole DNA extract was performed. The *G. m. morsitans* used here was a kind gift from Dr. Grace Murilla of Kenya Agricultural Research Institute – Trypanosome Research Centre (KARI-TRC) and different from the Yale strain sequenced. The PCR products were purified, cloned,

sequenced and compared to those in the database. In addition, the genes were translated and protein sequences compared to predicted proteins in the *Glossina* database.

### 3.9.1 Extraction of DNA

Whole DNA was extracted using alcohol precipitation method. The *G. m. morsitans* flies used were a kind gift from Dr. Grace Murilla of Kenya Agricultural Research Institute – Trypanosoma Research Centre (KARI-TRC). A single fly was homogenized in 50 µl of grinding buffer (Appendix 1) in a 1.5 ml microcentrifuge tube using a pestle then incubated at 68 °C for 30 min. Seven microlitres of 8M potassium acetate was added to the homogenate and incubated in ice for 30 min. Subsequently, the mixture was centrifuged at 20,000×g for 15 min, the supernatant transferred to a clean 1.5 ml microcentrifuge tube and mixed with 80 µl of isopropanol. This was incubated at room temperature for 5 min then centrifuged at 20,000×g for 15 min and the supernatant discarded. The pellet was finally washed with 200 µl of 70 % ethanol, air dried for 10 min and resuspended in 100 µl of nuclease free water.

### 3.9.2 Polymerase chain reaction

Polymerase chain reaction (PCR) was done using One *Taq*® Hot Start DNA polymerase (New England Biolabs, Inc. Massachusetts, USA) which has 3' to 5' exonuclease activity that increases the fidelity of amplification. The reaction mix consisted of 10 µl of 5X buffer, 1 µl of 10mM dNTPs, 1 µl of 10 mM of each primer, 0.25 µl of DNA polymerase, 2 µl of DNA and 34.75 µl of nuclease free water to give a total of 50 µl. Amplification was done using an AB biosystems 9700 thermal cycler and a reaction condition of 30 sec hot start at 94 °C, followed by 35 cycles at 94 °C for 30 sec (denaturation), Tm °C for 1 min (annealing), and 68 °C for $X_t$ min (extension) and a final extension at 68 °C for 5 min. Tm and $X_t$ values varied as shown in table 1.

**Table 1**. **List of genes, primer sequences, annealing temperatures (Tm), extension time (X$_t$) and target product size.** 'For' refers to the forward primer sequence, 'Rev' refers to the reverse primer sequence.

| Gene | Primer Sequence (5' to 3') | Annealing Temperature, Tm (°C) | X$_t$ (min) | Product Size (bp) |
|---|---|---|---|---|
| GMOY009051 | For: ATGAAACCTTTTAAGCAG | 47 | 1:20 | 1055 |
|  | Rev: TTAATCATCATCCGAAAA |  |  |  |
| GMOY009056 | For: ATGCTTAACTATTTACCG | 53 | 1:20 | 1602 |
|  | Rev: TTAAGTGCCTCCGTTAGA |  |  |  |
| GMOY005148 | For: ATGAGCAAGTTAAGCATTT | 52 | 1:20 | 1035 |
|  | Rev: TTAAGGTGCCAGACGTTCA |  |  |  |
| GMOY002354 | For: ATGACGCACTGGGAGGAC | 46 | 1:00 | 921 |
|  | Rev: CTAAATAAATTCGCTGTG |  |  |  |
| GMOYdhfr2 | For: ATGTTGAAATTTAATTTAAT | 46 | 1:00 | 624 |
|  | Rev: TTATTCTCGTTTTTCTAGTA |  |  |  |

Following amplification, the PCR products were gel electrophoresed on 1% agarose gel and products gel purified (see section 3.9.2 below).

### 3.9.3 Gel extraction of PCR products

After gel electrophoresis, the gel was viewed under UV-light and amplified DNA products were individually excised using sterile blades. Gel extraction was carried out using QuickClean II gel extraction kit (GeneScript, New Jersey, USA) as follows. The excised bands were weighed and excess gel cut out so as not to exceed 400 mg per test. One volume of gel slices were dissolved in three volumes of binding buffer in 1.5 ml microcentrifuge tube and dissolved by incubation at 50 °C in a water bath for about 10 min with occasional vortexing. One volume of isopropanol to 1 volume of gel was added to the dissolved gel. The mixture was transferred to a spin column and centrifuged at 6000×g for 1 min. Five hundred microlitres of binding buffer was added to the column and centrifuged again at 6000×g for 1 min and the flow through discarded. Subsequently, 750 µl of wash buffer was added to the column and left to stand for 5 min followed by centrifugation at 12000×g for 1 min. The column was centrifuged again at 12000×g for 1 min to completely remove the wash buffer. Lastly the column was transferred to a clean microcentrifuge tube and 25 µl of elution buffer added and left to stand for 5 min. The column was then centrifuged at 12000×g for 2 min and

the flow through which contain the amplification product retained for cloning. Before cloning, the success of the extraction was verified by gel electrophoresis of 3 µl of the eluted PCR product on a 1% ethidium bromide stained agarose gel.

### 3.9.4 Gene cloning using pGEM®-T Easy vector

The purified products of amplification were clonded into pGEM-T® easy expression system (Promega, Madison, USA) using manufacturer's instructions. Ligation mix of 5 µl of 2X rapid ligation buffer, 1 µl of 50 ng pGEM-T® easy vector, 3 µl of gel purified PCR product, 1 µl of T4 DNA ligase (3 weiss units/µl) was prepared and incubated overnight at 4 °C. Subsequently, high efficiency competent JM109 cells ($\geq 1 \times 10^8$ cfu/µg DNA) were transformed using the ligation mix as follows. For each reaction, a sterile 1.5 ml microcentrifuge tube was pre-chilled on ice and 50 µl of competent cells added. Two microlitres of ligation reaction mix were added and mixed gently by flicking before incubating in ice for 20 min. Transformation was carried out by heatshocking at 42 °C for 45 seconds in a water bath and immediately returned in ice for 2 min. 950 µl of SOC medium (appendix 2) was added and cells incubated for 1.5 hours at 37 °C with shaking (~150rpm). Afterwards, the cells were concentrated by centrifuging at 3000×g for 5 min and 100 µl of transformation culture retained for resuspension. The resuspended cells were plated on LB (Luria-Bertani) agar plates (appendix 2) containing ampicillin, Isopropyl $\beta$-D-1-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-$\beta$-D-galactopyranoside (x-Gal) (appendix 2) and incubated overnight at 37°C.

Single white colonies were subsequently picked from the LB agar plates and inoculated into 5 ml LB broth (appendix 2) at 37°C overnight in an incubator with shaking (~150rpm). These were used for plasmid purification.

### 3.9.5 Recombinant plasmid purification, sequencing and analysis

Plasmids from the transformed cells were purified using Quick Clean 5M miniprep kit (GeneScript, New Jersey, USA) using manufacturer's instructions. 1.5 ml of LB broth cultured cells were pelleted by centrifuging at 12000×g for 3 min. The pelleted cells were resuspended in 100 µl resuspension buffer by flicking the tube. 200 µl lysis buffer was added to the resuspended cells and gently mixed by inverting the tube 4-6 times. Subsequently, 300 µl of neutralization buffer was added and mixed by inverting the tube 4-6 times. The mixture was then centrifuged at 12000×g for 7 min and the supernatants transferred to spin columns. The spin columns were centrifuged at 12000×g for 1 min and the flow through discarded. 500 µl wash buffer was added to each column and centrifuged at 12000×g for 1 min. This wash

step was repeated. To completely remove the wash buffer, the spin column was centrifuged at the same condition for 1 min. Thereafter, the column was transferred to new sterile 1.5 ml microcentrifuge tubes, 25 µl of nuclease free water added and left to stand for 5 min before elution by centrifugation at 12000×g for 2 min. The recovered plasmids were electrophoresed and respective inserts amplified using their respective primers as in 3.9.1.

Sequencing of the insert was outsourced from Macrogen (Korea, Seoul, South Korea). Eventually, the sequence obtained from the PCR amplification was translated, then aligned with the sequence of the respective predicted gene models using ClustalW (Thompson *et al.*, 1994).

# CHAPTER FOUR
# RESULTS

## 4.1 Global vitamin and cofactor biosynthesis pathways

To enable accurate determination of *Glossina* B-vitamin and cofactor biosynthesis enzymes, all experimentally validated and computationally annotated enzymes involved in these processes were identified across the three kingdoms. A total of 144 experimentally validated orthologs that participate in B-vitamins and cofactors biosynthesis pathways were identified in eukaryotes and prokaryotes and used to search orthologs in *Glossina*. A table of these genes is provided in the supplementary material (Table S1). Subsequently, their physicochemical properties namely protein lengths, molecular weights (MW) and isoelectric points (pI), and the domain organization were determined (see Table S2 and Table S3 in the supplementary material). To have a global outlook of all the possible enzymes and pathways, the experimentally validated orthologs were used for searches and an additional 440 non-experimentally validated orthologs identified. This data together with documented biochemical pathways for various organisms were used to generate integrated pathways for biosynthesis of B-vitamins and cofactors (Figure 14). The pathways generally reveal that bacteria encode the highest number of enzymes for biosynthesis of most B-vitamins and cofactors while insects have the lowest number.

In thiamine biosynthesis pathway, only bacteria, plants and yeast have the full pathway. Insects and human do not synthesize thiamine but may acquire it from other sources, possibly diet in order to synthesize the cofactor thiamine pyrophosphate (TPP).

Bacteria, yeast and plants have a full riboflavin biosynthesis pathway. Insects and humans on the other hand do not biosynthesize riboflavin but synthesize the cofactor flavin adenine dinucleotide (FAD) from outsourced riboflavin. In nicotinamide adenine dinucleotide (NAD) biosynthesis pathway, plants and bacteria utilize the aspartate (*de novo*) or the nicotinate pathway while animals utilize the nicotinamide pathway (Lin *et al.*, 2010).

## A) Nicotinamide (vitamin B3)
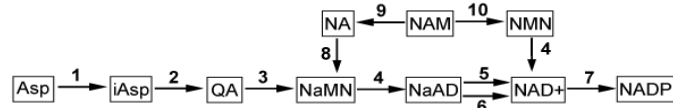


| Code | Enzyme | Insects | | | Bacteria | | Plants | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Dm | Ag | Tc | Ec | Bs | At | Sb | Sc | Hs |
| 1 | L-aspartate oxidase | - | - | - | + | + | + | + | - | - |
| 2 | Quinolinate synthase A | - | - | - | + | + | + | + | - | - |
| 3 | Quinolinate phosphoribosyl transferase | - | - | - | + | + | + | + | + | + |
| 4 | NMN adenylyltransferase | + | + | + | + | + | + | + | + | + |
| 5 | Glutamine dependent NAD synthetase | + | + | + | - | - | + | + | + | + |
| 6 | NH (3) dependent NAD synthetase | - | - | - | + | + | - | - | - | - |
| 7 | NAD kinase | + | + | + | + | + | + | + | - | + |
| 8 | Nicotinate phosphoribosyl transferase | + | + | + | + | + | + | + | + | + |
| 9 | Nicotinamidase | - | - | - | + | + | - | - | + | - |
| 10 | Nicotinamide phosphoribosyltransferase | - | - | - | - | - | - | - | - | + |

KEY
| | | | |
|---|---|---|---|
| Asp | L-aspartate | NAM | Nicotinamide |
| iAsp | Iminoaspartate | NaMN | Nicotinatemononucleotide |
| QA | Quinolinate | NaAD | Nicotinate adenine dinucleotide |
| NADP | NAD phosphate | NAD | Nicotinamide adenine dinucleotide |
| NA | Nicotinic acid | NMN | Nicotinamide mononucleotide |

## B) Pantothenate (vitamin B5)



| Code | Enzyme | Insects | | | Bacteria | | Plants | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Dm | Ag | Tc | Ec | Bs | At | Sb | Sc | Hs |
| 1 | Branched-chain-amino-acid aminotransferase | + | + | + | + | + | + | + | + | + |
| 2 | Ketopantoate hydroxymethyltransferase | - | - | - | + | + | + | + | + | - |
| 3 | Ketopantoate reductase | - | - | - | + | + | - | - | + | + |
| 4 | Dihydropyrimidine dehydrogenase | + | + | + | + | + | + | - | - | + |
| 5 | Dihydropyrimidinase | + | + | + | + | + | + | + | + | + |
| 6 | Beta-ureidopropionase | + | + | + | + | + | - | + | - | + |
| 7 | Aspartate 1-decarboxylase | - | - | - | + | + | - | - | - | - |
| 8 | Pantothenate synthetase | - | - | - | + | + | + | + | + | + |
| 9 | Pantothenate kinase | + | + | + | + | + | + | + | + | + |
| 10 | Phosphopantothenate-cysteine ligase | + | + | + | + | + | + | + | + | + |
| 11 | Phosphopantothenoylcysteine decarboxylase | + | + | + | + | + | + | + | + | + |
| 12 | Phosphopantetheine adenylyltransferase | + | + | + | + | + | + | + | + | + |
| 13 | Dephospho-CoA kinase | + | + | + | + | + | + | + | + | + |

KEY
| | | | | | |
|---|---|---|---|---|---|
| Val | L- valine | Asp | L- aspartate | CAla | N-carbamoyl β-alanine |
| Ura | Uracil | dUra | 5,6-dihydro-uracil | pPan | 4'-phosphopantothenate |
| Ala | β-alanine | CoA | Coenzyme A | pPanCy | 4'-phosphopantothenoyl cysteine |
| KPa | α- ketopantoate | Pan | Pantothenate | pPanN | 4'-phosphopantothenine |
| Pa | Pantoate | KiV | α- ketoisovalerate | DCoA | Dephospho coenzyme A |

## C) Thiamine (vitamin B1)



| Code | Enzyme | Insects | | | Bacteria | | Plants | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Dm | Ag | Tc | Ec | Bs | At | Sb | Sc | Hs |
| 1 | Cysteine desulfurase | + | + | + | + | + | + | + | + | + |
| 2 | Thiazole synthase | - | - | - | + | + | - | - | - | - |
| 3 | Thiamine biosynthesis protein ThiH | - | - | - | + | - | - | - | - | - |
| 4 | Glycine oxidase | - | - | - | + | - | - | - | - | - |
| 5 | Phosphomethylpyrimidine synthase | - | - | - | + | + | + | + | - | - |
| 6 | Phosphomethylpyrimidine kinase | - | - | - | + | + | + | + | - | - |
| 7 | Thiamine-phosphate synthase | - | - | - | + | + | + | + | + | - |
| 8 | Thiamine-monophosphate kinase | - | - | - | + | + | - | - | - | - |
| 9 | Non-specific phosphatase | + | + | + | - | + | + | + | + | + |
| 10 | Thiamine pyrophosphokinase | + | + | + | - | + | + | + | + | + |
| 11 | Hydroxyethylthiazole kinase | - | - | - | + | + | + | + | + | - |

KEY
| | | | |
|---|---|---|---|
| Dx | 1-deoxy D-xylulose 5-phosphate | HMPP | Hydroxymethyl pyrimidinepyrophosphate |
| dHG | Dehydroglycine | TMP | Thiamine monophosphate |
| TPC | Thiazole phosphate carboxylate | TPP | Thiamine pyrophosphate |
| Az | 5-aminoimidazole | TTP | Thiamine triphosphate |
| HMP | Hydroxymethyl pyrimidine phosphate | HMT | Hydroxyethyl methylthiazole |

## D) Pyridoxine (vitamin B6)



| Code | Enzyme | Insects | | | Bacteria | | Plants | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Dm | Ag | Tc | Ec | Bs | At | Sb | Sc | Hs |
| 1 | D-erythrose-4-phosphate dehydrogenase | - | - | - | + | - | - | - | - | - |
| 2 | Erythronate-4-phosphate dehydrogenase | - | - | - | + | - | - | - | - | - |
| 3 | Phosphoserine aminotransferase | + | + | + | + | + | + | - | + | + |
| 4 | 4-hydroxythreonine-4-phosphate dehydrogenase | - | - | - | + | - | - | - | - | - |
| 5 | Pyridoxine 5'-phosphate synthase | - | - | - | + | + | - | - | - | - |
| 6 | Pyridoxine-5'-phosphate oxidase | + | + | + | + | - | - | - | + | + |
| 7 | Pyridoxal phosphate phosphatase | + | + | + | - | - | - | - | - | + |
| 8 | Pyridoxal kinase | + | + | + | + | + | + | - | + | + |

KEY
| | | | |
|---|---|---|---|
| EP | D-erythrose 4-phosphate | PLP | Pyridoxal phosphate |
| PEA | 4-phospho D- erythronate | PMP | Pyridoxamine monophosphate |
| Pxm | Pyridoxamine | PHT | 4-phosphohydroxy L-threonine |
| Pxn | Pyridoxine | dXP | 1-deoxyxylulose 5-phosphate |
| PNP | Pyridoxine phosphate | HPKB | 3-hydroxy 4-phosphohydroxy α-ketobutyrate |

## E) Folate (vitamin B9)



| Code | Enzyme | Distribution | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Insects | | | Bacteria | | Plants | | | |
| | | Dm | Ag | Tc | Ec | Bs | At | Sb | Sc | Hs |
| 1 | GTP cyclohydrolase I | + | + | + | + | + | + | - | + | + |
| 2 | Alkaline phosphatase | + | + | + | + | + | + | + | + | + |
| 3 | Dihydroneopterin pyrophosphatase | - | - | - | + | - | + | - | - | - |
| 4 | Non specific phosphatase | + | + | + | + | + | + | + | + | + |
| 5 | Dihydroneopterin aldolase | - | - | - | + | + | + | + | + | - |
| 6 | 2-amino-4-hydroxy-6-hydroxymethyl dihydropteridine pyrophosphokinase | - | - | - | + | + | + | + | + | - |
| 7 | P-aminobenzoic acid synthase | - | - | - | + | + | + | + | + | - |
| 8 | Aminodeoxychorismate lyase | - | - | - | + | + | - | - | - | - |
| 9 | Dihydropteroate synthase | - | - | - | + | + | + | + | + | - |
| 10 | Folylpolyglutamate synthase | + | + | + | + | + | + | + | + | + |
| 11 | Dihydrofolate reductase | + | - | + | + | + | + | + | + | + |

**KEY**

| | |
|---|---|
| GTP | Guanosine triphosphate |
| dHPTP | Dihydroneopterin triphosphate |
| dHT | Dihydroneopterin |
| dHPMP | Dihydroneopterin monophosphate |
| HMdHT | Hydroxymethyldihydropterin |
| HMdHTP | Hydroxymethyldihydropterin pyrophosphate |

| | |
|---|---|
| CM | Chorismate |
| ADCM | Aminodeoxychorismate |
| PABA | Para aminobenzoic acid |
| dHTe | Dihydropteroate |
| dHf | Dihydrofolate |
| tHF | Tetrahydrofolate |

## F) Riboflavin (vitamin B2)



| Code | Enzyme | Distribution | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Insects | | | Bacteria | | Plants | | | |
| | | Dm | Ag | Tc | Ec | Bs | At | Sb | Sc | Hs |
| 1 | GTP cyclohydrolase-2 | - | - | - | + | + | + | + | + | - |
| 2 | DRAP deaminase | - | - | - | + | + | + | + | - | - |
| 3 | APAsy uracil reductase | - | - | - | + | + | + | + | - | - |
| 4 | DARP reductase | - | - | - | - | - | - | - | + | - |
| 5 | Bifunctional protein RIB2 | - | - | - | - | - | - | - | + | - |
| 6 | Lumazine synthase | - | - | - | + | + | + | + | + | - |
| 7 | DHBP synthase | - | - | - | + | + | + | + | + | - |
| 8 | Riboflavin synthase | - | - | - | + | + | + | - | + | - |
| 9 | Riboflavin kinase | + | - | + | + | + | + | + | + | + |
| 10 | FAD synthase | + | + | + | + | + | - | - | + | + |

**KEY**

| | |
|---|---|
| GTP | Guanosine 5'triphosphate |
| ARAP | 5-amino 6-ribitylamino 2,4-(1H,3H)-pyrimidinedione |
| APAsyU | 5-amino 6-(5-phosphoD-ribosyl amino) uracil |
| APAtyU | 5-amino 6-(5-phospho D-ribityl amino) uracil |
| DRAP | 2,5-diamino 6-ribosylamino 4-(3H)-pyrimidinone 5'phosphate |

| | |
|---|---|
| RP | Ribulose 5phosphate |
| DHBP | 3,4-dihydroxy 2-butanone 4-phosphate |
| dRL | 6,7-dimethyl 8-ribityl lumazine |
| RF | Riboflavin |
| FMN | Flavin mononucleotide |
| FAD | Flavin adenine dinucleotide |

## G) Biotin (vitamin B7)



| Code | Enzyme | Distribution | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Insects | | | Bacteria | | Plants | | | |
| | | Dm | Ag | Tc | Ec | Bs | At | Sb | Sc | Hs |
| 1 | 8-amino-7-oxononanoate synthase | - | - | - | + | + | + | + | - | - |
| 2 | Adenosylmethionine-8-amino-7-oxononanoate aminotransferase | - | - | - | + | + | + | + | + | - |
| 3 | Dethiobiotin synthetase | - | - | - | + | + | - | - | + | - |
| 4 | Biotin synthase | - | - | - | + | + | + | - | + | - |
| 5 | Biotin protein ligase | + | + | + | + | + | - | - | + | + |

**KEY**

| | |
|---|---|
| PT | Pimelate thioester |
| Aon | 8-amino-7-oxononanoate |
| Don | 7,8-diaminononanoate |
| dTB | Dethiobiotin |
| Biotin-P | Biotin Protein |

**Figure 14. Integrated pathways for vitamins and cofactor biosynthesis across insects, bacteria, plants, yeast and man.** The substrates/ products are represented in boxes and their abbreviations are shown in the KEY. Arrows represent chemical reactions and numbers represent enzymes. In the tables, plus (+) and minus (-) signs indicate presence and absence of an enzyme respectively. Abbreviations: Dm, *Drosophila melanogaster*; Ag, *Anopheles gambiae*; Tc, *Tribolium castenum*; Ec, *Escherichia coli*; Bs, *Bacillus subtilis*; At, *Arabidopsis thaliana*; Sb, *Sorghum bicolor*; Sc, *Saccharomyces cerevisiae*; Hs, *Homo sapien*.

In pantothenate biosynthesis pathway, insects and humans do not have the ability to synthesize pantoate and condense pantoate with alanine to make pantothenate (vitamins B5). All other organisms analyzed have enzymes for synthesizing coenzyme A (CoA) from vitamin B5.

*Escherichia coli* and plants are the only organisms that biosynthesize pyridoxine (vitamins B6) *de novo* (Tambasco-Studart *et al.*, 2005). Other organisms salvage pyridoxal phosphate (PLP) from vitamin B6 or from the ketone or amine forms of the vitamin i.e. pyridoxal and pyridoxamine respectively.

Bacteria, plants and yeast have a complete tetrahydrofolate (THF) biosynthesis pathway. Insects and humans on the other hand do not have the potential to biosynthesize para amino benzoic acid (PABA), a major component of THF, and therefore do not synthesize the cofactor *de novo*. In biotin biosynthesis pathway, only bacteria and yeast have the potential to synthesize biotin *de novo* but plants pathway is not well known (Pinon *et al.*, 2005). The other organisms only have biotin protein ligase, which is the enzyme that ligates biotin to the biotin dependent enzyme.

In summary, only bacteria, plants and yeast have the potential to biosynthesize B-vitamins while insects and humans lack the *de novo* biosynthesis pathways. Most organisms however have the potential to biosynthesize cofactors from the B-vitamins.

## 4.2 Global domain organization in B-vitamins and cofactors biosynthesis enzymes

The domain organization of enzymes involved in B-vitamins and cofactors biosynthesis is shown in figure 15. Generally the domain architecture for orthologous enzymes across insects, bacteria, plants, yeast and human varies. It is only in six enzymes namely, cysteine desulfurase, thiamine pyrophosphokinase, phosphoserine aminotransferase, nicotinamide mononucleotide adenyltransferase, NAD kinase and branched chain amino acid aminotransferase, that all the organisms have similar domain architecture.

Among insects, the architecture is similar for most enzymes except for nicotinate phosphoribosyltransferase (in *Tribolium* and *Anopheles*), dihydropyrimidine dehydrogenase in *Tribolium*) and phosphopantothenate cysteine ligase (in *Drosophila* and *Tribolium*). Most of the difference in bacteria domains was observed between the free living and the endosymbionts whereas in plants, *Arabidopsis* and *Sorghum* have similar domain organization in most instances.

| Pathway | Enzyme | Domain variants |
|---|---|---|
| Thiamine pyrophosphate/ Vitamin B1 | Cysteine desulfurase | All |
| | Sulfur carrier protein ThiS | Bacteria |
| | Thiazole biosynthesis adenylyltransferase ThiF | Bacteria |
| | Thiamine biosynthesis protein ThiH | Bacteria; *Sodalis* |
| | Glycine oxidase | *Bacillus* |
| | Hydroxyethylthiazole kinase | Plants, Bacteria; Yeast |
| | Thiazole synthase | Bacteria |
| | Hydroxymethylpyrimidine phosphate synthase | Bacteria, Plants |
| | Phosphomethylpyrimidine kinase | Plants; Yeast; Bacteria |
| | Thiamine-phosphate synthase | Plants; Yeast; Bacteria |
| | Thiamine-monophosphate kinase | Bacteria |
| | Thiamine pyrophosphokinase | All |
| Flavin adenine dinucleotide/ Vitamin B2 | 3,4-dihydroxy-2-butanone 4-phosphate synthase | Plants, *Bacillus*; Bacteria, Yeast |
| | GTP cyclohydrolase-2 | Plants, *Bacillus*; Bacteria, Yeast |
| | DARP reductase | Yeast |
| | Bifunctional protein RIB2 | Yeast |
| | DRAP deaminase | Bacteria; Plants |
| | HTP reductase | Bacteria |
| | Lumazine synthase | Bacteria, Yeast, Plants |
| | Riboflavin synthase | Bacteria, Yeast, Plants |
| | Riboflavin kinase | *Drosophila, Tribolium*, Human; Bacteria |
| | FAD synthase | *Drosophila, Anopheles*, Yeast; *Tribolium*, Human; Bacteria |
| Pyridoxal phosphate/ Vitamin B6 | D-erythrose-4-phosphate dehydrogenase | *Escherichia* |
| | Erythronate-4-phosphate dehydrogenase | Bacteria |
| | Phosphoserine aminotransferase | All |
| | 4-hydroxythreonine-4-phosphate dehydrogenase | Bacteria |
| | Pyridoxine 5'-phosphate synthase | Bacteria |
| | Pyridoxal kinase | Drosophila, Bacteria, Yeast; *Tribolim, Anopheles*, Human |
| | PLP phosphatase | Insects, Human |
| | Pyridoxine-5'-phosphate oxidase | Insects, Bacteria, Human |

Legend:

| | | | | | |
|---|---|---|---|---|---|
| Aminotran_5 | ThiG | AIRS | PseudoU_synth_2 | MoCF_biosynth | PdxJ |
| ThiS | ThiC | AIRS_C | dCMP_cyt_deam_1 | Gp_dh_N | Phos_pyr_kin |
| ThiF | ThiC_associated | TPK_catalytic | DMRL_synthase | Gp_dh_C | PfkB |
| MoeZ_MoeB | Phos_pyr_kin | TPK_B1_binding | Lum_binding | 2-Hacid_dh | Put_Phosphatase |
| Radical_SAM | TMP-TENI | DHBP_synthase | Flavokinase | 2-Hacid_dh_C | Pyridox_oxidase |
| BATS | TENA_THI-4 | GTP_cyclohydro2 | FAD_syn | DUF3410 | PNPOx_C |
| DAO | HK | RibD_C | PAPS_reduct | PdxA | SerC |

200 aa

41

| Pathway | Enzyme | Domain variants |
|---|---|---|
| Nicotinamide adenine dinucleotide/ Vitamin B3 | L-aspartate oxidase | Bacteria, Plants |
| | Quinolinate synthase A | Bacteria, *Sorghum* / *Arabidopsis* |
| | Nicotinamide phosphoribosyltransferase | Human |
| | Nicotinamidase | Insects / Bacteria, Yeast |
| | Nicotinate phosphoribosyltransferase | *Glossina, Drosophila*, Bacteria, Yeast, Human / *Anopheles, Tribolium*, Plants |
| | Nicotinamide mononucleotide adenylyltransferase | All |
| | NAD synthetase | Insects, Human, Yeast, Plants / Bacteria |
| | NAD kinase | All |
| | Quinolinate phosphoribosyltransferase | Bacteria, Plants, Yeast, Human |
| Co-enzyme A/ Vitamin B5 | Branched chain amino acid aminotransferase | All |
| | Ketopantoate hydroxymethyltransferase | Bacteria, Plants, Yeast |
| | Ketopantoate reductase | Bacteria, Yeast |
| | Dihydropyrimidine dehydrogenase | *Glossina, Drosophila*, Human / *Tribolium* / Plants |
| | Dihydropyrimidinase | Insects, Human / Bacteria |
| | Beta-ureidopropionase | Insects, Plants, Human |
| | Aspartate 1-decarboxylase | Bacteria |
| | Pantothenate synthetase | Bacteria, Plants |
| | Pantothenate kinase | Insects / Bacteria / Insects / Bacillus / Plants, Human |
| | Phosphopantothenate-cysteine ligase | *Glossina, Tribolium, Arabidopsis*, Yeast / *Drosophilla, Anopheles, Sorghum*, Human / Bacteria |
| | Phosphopantothenoylcysteine decarboxylase | Insects, Plants, Yeast, Human / Bacteria |
| | Phosphopantetheine adenylyltransferase | Insects, Human / Bacteria, Plants, Yeast |
| | Dephospho-CoA kinase | Insects, Human / Bacteria, Plants, Yeast |

Legend:
FAD_binding_2, Succ_DH_flav_C, NadA, SufE, NAPRTase, EF-hand_7, Isochorismatase, CTP_transf_2, CN_hydrolase, NAD_synthase, QRPTase_N, QRPTase_C, Aminotran_4, Pantoate_transf, ApbA, ApbA_C, Fer4_20, NAD_binding_8, Asp_decarbox, DHO_dh, Fer4_21, CN_hydrolase, Amidohydro_1, Amidohydro_4, CoaE, Pantoate_ligase, Fumble, DUF89, PRK, Pan_kinase, DFP, Flavoprotein, CTP_transf_2, 200 aa

42
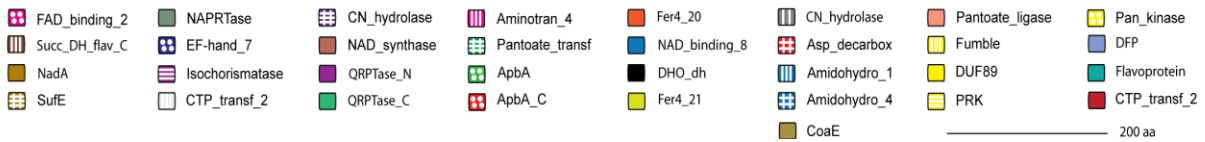
**Figure 15. Domain architecture for orthologs involved in B-vitamins and cofactors biosynthesis pathways.** The domain architecture of experimentally validated and computer predicted orthologs was determined at public protein databases namely interpro, pfam and prosite. The different colour codes represent different domains and their names are given below the figure. The enzyme names in blue are enzymes for cofactors biosynthesis while the rest in black are for B-vitamins biosynthesis

The variation in domain organization can be due to loss of and/or acquisition of domains. For example, thiamine biosynthesis protein H (*Sodalis*), DRAP deaminase (plants), riboflavin kinase (human and insects), dihydropyrimidine dehydrogenase (*Tribolium*), NAD synthetase

(bacteria) and biotin protein ligase (*Wiggleswothia* and *Wolbachia*) lack BATS, RibD_C, FAD_syn, NAD_binding_8, CN_hydrolase and HTH_11 and BPL_C domains respectively, domains that are present in orthologs. In addition, the domains for folylpolyglutamate synthase (*Escherichia* and *Bacillus*) and dethiobiotin synthetase (*Sodalis*) have been lost in the other organisms. Some organisms may have acquired extra domains like in FAD synthase (*Tribolium* and human), quinolinate synthase A (*Arabidopsis*) and nicotinamidase (insects), (Figure 15).

Another observation was presence of multiple domain proteins suggesting multifunctional enzymes. Examples are hydroxyethylthiazole kinase, phosphomethylpyrimidine kinase and thiamine phosphate synthase (vitamin B1 pathway), 3,4-dihydroxy 2-butanone 4-phosphate synthase, GTP cyclohydrolase-2, bifunctional protein RIB2, riboflavin kinase and FAD synthase (vitamin B2 pathway), dihydropyrimidine dehydrogenase (vitamin B5 pathway), all CoA biosynthesis enzymes, dihydroneopterin aldolase, 7,8-dihydro 6-hydroxymethylpterin pyrophosphokinase, PABA synthase and dihydropteroate synthase (vitamin B9 pathway), dihydrofolate reductase (THF pathway), DAPA aminotransferase and dethiobiotin synthetase (biotin pathway) (Figure 15). For instance in yeast, hydroxyethylthiazole kinase has two domains, one for hydroxyethylthiazole kinase activity (HK) and the other for thiamine phosphate synthase activity (TMP-TENI), indicating two enzyme activities on a single polypeptide.

In addition, some multifunctional enzymes namely hydroxymethylthiazole kinase, phosphomethylpyrimidine kinase, thiamine phosphate synthase (vitamin B1 pathway), dihydroneopterin aldolase, 7,8-dihydro 6-hydroxymethylpterin pyrophosphokinase and dihydropteroate synthase (vitamin B9 pathway), have different combination of domains and therefore different combination of enzymes in a single polypeptide. For instance in thiamine biosynthesis, plants combine phosphomethylpyrimidine kinase and thiamine phosphate synthase while yeast combines phosphomethylpyrimidine kinase and hydroxymethylpyrimidine phosphate synthase. Lastly some organisms have multiple domains for some enzymes namely GTP cyclohydrolase I (*Arabidopsis*), dihydroneopterin aldolase (yeast) and nicotinate phosphoribosyltransferase (*Anopheles*, *Tribolium* and plants).

In summary, domain architecture of B-vitamins and cofactors biosynthesis enzymes vary widely across organisms analyzed. This diversity is present within and between groups of organisms. Variations include loss of domains that most likely results into specialized

enzymes, or domain acquisitions and fusion of enzymes that result in a single polypeptide with multiple different enzymatic activity i.e. multifunctional polypeptides.

**4.3 *Glossina* B-vitamins and cofactors biosynthesis pathway**

To identify genes in *Glossina* genome that encode enzymes in B-vitamins and cofactors biosynthesis pathways, all experimentally validated and the computationally validated orthologs used to design the global pathways (Figures 14 and 15) were used as query sequences for homology search at VectorBase (www.vectorbase.org). Both tblastn and blastp were used respectively to search the nucleotide sequence and the software predicted protein models of the *Glossina* genome. All the hits obtained in multiple searches and which had low e-values were analysed for start and stop codons, and splice sites then translated to proteins. Both tblastn and blastp searches resulted in the same protein and these protein sequences were used in subsequent analyses with numerical IDs bearing the prefix GMOY, which stands for *Glossina morsitans* Yale strain. In addition, their physicochemical properties and domain organization were analysed and the proteins that resembled the experimentally validated orthologs in these properties were used in a reverse blastp search and only hits that identified the initial query protein were considered as true hits.

A total of 34 genes encoding proteins involved in vitamins and cofactors biosynthesis were identified in *Glossina* (Table 2). Six of the 34 genes encode enzymes involved in biosynthesis of B-vitamins namely thiamine (1 enzyme), pyridoxine (1 enzyme) and pantothenate (4 enzyme) (Table 2a), while the remaining 28 genes encode for 20 enzymes in cofactors biosynthesis namely TPP (1 enzyme), FAD (2 enzymes), NAD (5 enzymes), CoA (5 enzymes), PLP (3 enzymes) and THF (4 enzymes) (Table 2b). The variation in genes and enzyme numbers i.e. 28 genes vs 20 enzymes, is due to multiple genes encoding a single enzyme protein as in the case of FAD synthase (2 gene copies), pantothenate kinase (2 gene copies), dihydrofolate reductase (2 gene copies) and alkaline phosphatase (6 gene copies) (Table 2b) and were considered isoenzymes (Appendix 5). Two genes namely GMOY000849 and GMOY009677 were allocated EC number based on the ortholog they cluster with in the phylogenetic trees shown in supplementary data (Figure S1). Together, the tsetse fly encodes 26 enzymes in B-vitamins and cofactors biosynthesis pathways.

**Table 2a. List of *Glossina* genes and their respective enzymes involved in B-vitamins biosynthesis**

| Pathway | Enzyme EC Number | Enzyme | Gene ID | Protein Length (aa) |
|---|---|---|---|---|
| Thiamine | EC:2.8.1.7 | Cysteine desulfurase | GMOY000270 | 467 |
| Pyridoxine | EC:2.6.1.52 | Phosphoserine aminotransferase | GMOY006775 | 364 |
| Pantothenate | EC:2.6.1.42 | Branched-chain-amino-acid aminotransferase | GMOY012027 | 450 |
| | EC:1.3.5.2 | Dihydropyrimidine dehydrogenase | GMOY000435 | 1035 |
| | EC:3.5.2.2 | Dihydropyrimidinase | GMOY004714 | 594 |
| | EC:3.5.1.6 | Beta-ureidopropionase | GMOY006358 | 386 |

**Table 2b. List of *Glossina* genes and their respective enzymes involved in cofactors biosynthesis**

| Pathway | Enzyme EC Number | Enzyme | Gene ID | Protein Length (aa) |
|---|---|---|---|---|
| TPP | EC:2.7.6.2 | Thiamine pyrophosphokinase | GMOY009051 | 291 |
| FAD | EC:2.7.1.26 | Riboflavin kinase | GMOY009056 | 165 |
| | EC:2.7.7.2 | FAD synthase | GMOY000849 | 424 |
| | EC:2.7.7.2 | FAD synthase | GMOY008278 | 247 |
| PLP | EC:2.7.1.35 | Pyridoxal kinase | GMOY003664 | 304 |
| | EC:3.1.3.74 | Pyridoxal phosphate phosphatase | GMOY009677 | 275 |
| | EC:1.4.3.5 | Pyridoxine-5'-phosphate oxidase | GMOY005148 | 254 |
| NAD | EC:3.5.1.19 | Nicotinamidase | GMOY006570 | 357 |
| | EC:2.4.2.11 | Nicotinate phosphoribosyltransferase | GMOY008976 | 480 |
| | EC:2.7.7.18; EC:2.7.7.1 | Nicotinamide mononucleotide adenylyltransferase 1 | GMOY002776 | 359 |
| | EC:6.3.5.1 | glutamine-dependent NAD(+) synthetase | GMOY008198 | 865 |
| | EC:2.7.1.23 | NAD kinase | GMOY010621 | 544 |
| CoA | EC:2.7.1.33 | Pantothenate kinase | GMOY006067 | 391 |
| | EC:2.7.1.33 | Pantothenate kinase | GMOY000935 | 495 |
| | EC:6.3.2.5 | Phosphopantothenate--cysteine ligase | GMOY002354 | 306 |
| | EC:4.1.1.36 | Phosphopantothenoylcysteine decarboxylase | GMOY009241 | 189 |
| | EC:2.7.1.24; EC:2.7.7.3 | Bifunctional coenzyme A synthase | GMOY000596 | 553 |
| | EC:2.7.1.24 | Dephospho-CoA kinase | GMOY006891 | 240 |
| THF | EC:3.5.4.16 | GTP cyclohydrolase 1 | GMOY009349 | 384 |
| | EC:3.1.3.1 | Alkaline phosphatase | GMOY004796 | 539 |
| | EC:3.1.3.1 | Alkaline phosphatase | GMOY000067 | 547 |
| | EC:3.1.3.1 | Alkaline phosphatase | GMOY009926 | 526 |
| | EC:3.1.3.1 | Alkaline phosphatase | GMOY006875 | 541 |
| | EC:3.1.3.1 | Alkaline phosphatase | GMOY004885 | 589 |
| | EC:3.1.3.1 | Alkaline phosphatase | GMOY007323 | 611 |
| | EC:6.3.2.17 | Folylpolyglutamate synthase | GMOY005468 | 524 |
| | EC:1.5.1.3 | Dihydrofolate reductase | GMOY008444 | 343 |
| | EC:1.5.1.3 | Dihydrofolate reductase | GMOYdhfr2 | 187 |
| Biotin | EC:6.3.4.10 | Biotin protein ligase | GMOY010708 | 1141 |

Abbreviations: TPP; thiamine pyrophosphate, FAD; flavin adenine dinucleotiede, NAD; nicotinamide adenine dinucleotide, CoA; coenzyme A, PLP; pyridoxal phosphate, THF; tetrahydrofolate. GMOYdhfr2 is the second gene encoding dihydrofolate reductase (dhfr).

**4.4 Domain structure of *Glossina* enzymes**

The *Glossina* enzymes share orthology with other organisms' enzymes for B-vitamins and cofactors biosynthesis, and this is confirmed further by similar domains organizations though with some variations as in the global analysis (Figures 15 and 16). *Glossina* cysteine desulfurase, thiamine pyrophosphokinase, phosphoserine aminotransferase, nicotinamide mononucleotide adenyltransferase, NAD kinase and branched chain amino acid aminotransferase (Figure 16) have same domain architecture as other eukaryotes (Figure 15). For example yeast, human and *Drosophila* thiamine pyrophosphokinase have TPK_catalytic and TPK_B1_binding domains at the N and C-terminals respectively, an organization present in *Glossina* thiamine pyrophosphokinase.

Some enzymes however exhibit variation in domain organization as observed in global domain analysis. For example, *Glossina* dihydrofolate reductase, bifunctional coenzyme A synthase and alkaline phosphatase domain organization (Figure 16) are different from those of insects, bacteria, plants, human and yeast (Figure 15). The rest of the enzymes in *Glossina* however have similar domain architecture as those of other insects.

*Glossina* has unique bifunctional coenzyme A synthase. In insects and humans, this enzyme consists of two domains CTP_transf_2 and CoaE for phosphopantetheine adenylyltransferase activity and dephospho-CoA kinase activity respectively in a single polypeptide i.e. bifunctional. However, in bacteria, yeast and plants, these two domains are encoded in separate genes hence independent polypeptides (Figure 15). Interestingly, *Glossina* encodes both the single and bifunctional enzymes (Figure 16). The gene encoding the bifunctional enzyme may have resulted from fusion of phosphopantetheine adenylyltransferase and dephospho-CoA kinase genes.

In addition, *Glossina* dihydrofolate reductase (DHFR) has a heat shock protein (HSP) domain which is absent in other orthologs. This HSP70 domain is absent in a second DHFR in *Glossina* that is similar to insects, yeast, human and bacteria DHFR (Figures 15 and 16) and the former *Glossina* DHFR is possibly due to domain acquisition. Despite the HSP70 domain, both *Glossina* polypeptides were assumed to be DHFR.

**Figure 16. Domain architecture of *Glossina* enzymes for B-vitamins and cofactors biosynthesis.** The domains for B-vitamins (A) and cofactors (B) biosynthesis enzymes was determined at public protein databases. The different shades represent different domains and their names are given below the figure. The IDs of genes are shown; * indicate enzymes with unique domains.

Though the domain for phosphoserine aminotransferase is similar across all organisms, in *Glossina* this enzyme has a FGGY motif at its C-terminal end, a motif observed in carbohydrate kinases. FGGY family of carbohydrate kinases include enzymes like L-fuculokinase, gluconokinase, glycerokinase, xylulokinase and L-xylulose kinase (Reizer *et al.*, 1991). It is therefore unique for this aminotransferase to contain a carbohydrate kinase motif.

Lastly, one of the six alkaline phosphatase isoenzymes encoded by gene GMOY004796 contains a phosphotransferase system (PTS) histidine-containing phosphocarrier protein (HPr) domain at its C-terminal in addition to the catalytic alkaline phosphatase domain. In contrast, all other alkaline phosphatases in *Glossina* have similar domain architecture like other homologs in insects.

In summary, the 26 enzymes obtained in *Glossina* have similar domain architecture to the other insects observed in the global analysis. Though some enzymes had unique domains, this characteristic is also present in other insects. With the *Glossina* genes having orthologs and the domain organization and physicochemical properties of their proteins similar to other organisms including insects, the genes were considered true and the resulting enzymes possibly functional in B-vitamins and cofactors biosynthesis.

**4.5 Gene validation**

To confirm that the genes annotated are truly existent in the fly, a polymerase chain reaction (PCR) was done to amplify full genes, one for each pathway. The shortest genes were selected for amplification because of their ease to amplify (Figure 17 and Table 1). Following amplification the products obtained were of the expected sizes (Figure 17) and after sequencing, the alignments between the protein sequences of the cloned genes and of the corresponding annotated enzymes were 98% identical (Table A1).
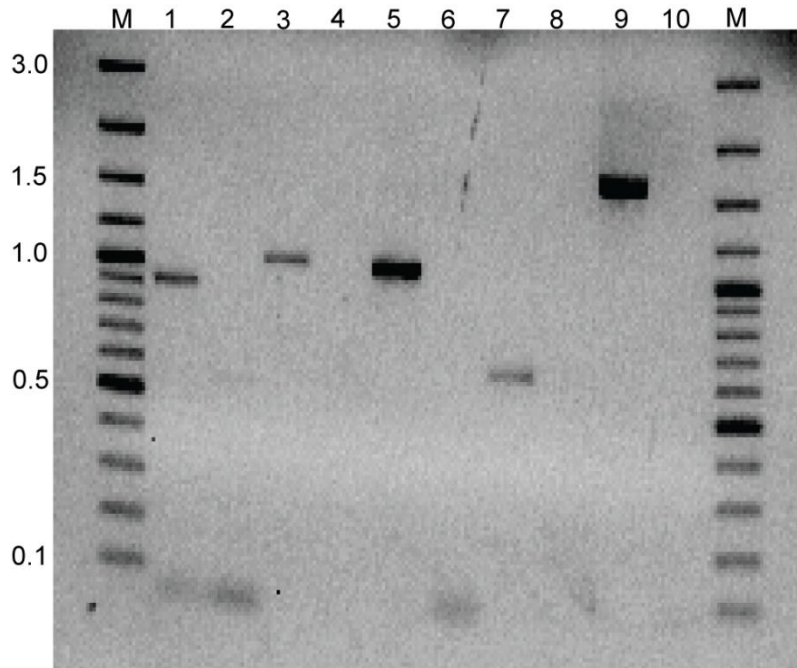
**Figure 17. Agarose gel image showing DNA bands of amplified genes.** The image was obtained after electrophoresis of PCR products of target genes on a 1% ethidium bromide stained agarose gel. Lane M represent the DNA ladder in kilobases while the subsequent lanes marked from 1-10 represent a gene and its negative control in the order GMOY002354, GMOY009051, GMOY005148, GMOYdhfr2 and GMOY009056.

## 4.6 Comparison between *Glossina* B-vitamins and cofactors biosynthesis enzymes and their orthologs

For comparative analysis, orthologs from a hematophagous insect, *An. gambiae* (also feeds on plant sap) *D. melanogaster* (feeds on ripe fruits) and *T. castanum* (flour) were used. Further, since endosymbionts of *Glossina* have been suggested to take part in B-vitamin metabolism (Pais *et al.*, 2008), *W. glossinidia*, *S. glossinidius*, *W. pipientis*, as well as free-living bacteria *E. coli* and *B. Subtilis* were included. The plants *A. thaliana* and *S. bicolor* which are free living and therefore biosynthesize B-vitamins, and model eukaryotes *H. sapiens* (humans) and *S. cerevisiae* (yeast) were also included. Figure 18 shows the comparative analysis.

| Pathway | Enzyme | Insects | Endosymbionts | Free-living bacteria | Plants | Yeast/Man |
|---|---|---|---|---|---|---|
| | | Gm, Ag, Dm, Tc | Wg, Wp, Sg | Ec, Bs | At, Sb | Sc, Hs |
| Thiamine pyrophosphate (Vitamin B1) | Cysteine desulfurase | | | | | |
| | Sulfur carrier protein ThiS | | | | | |
| | Thiazole biosynthesis adenylyltransferase ThiF | | | | | |
| | Thiamine biosynthesis protein ThiH | | | | | |
| | Glycine oxidase | | | | | |
| | Hydroxyethylthiazole kinase | | | | | |
| | Thiazole synthase | | | | | |
| | Hydroxymethylpyrimidine phosphate synthase | | | | | |
| | Phosphomethylpyrimidine kinase | | | | | |
| | Thiamine-phosphate synthase | | | | | |
| | Thiamine-monophosphate kinase | | | | | |
| | Thiamine pyrophosphokinase | | # | | | |
| Flavin adenine dinucleotide (Vitamin B2) | 3,4-dihydroxy-2-butanone 4-phosphate synthase | | | | | |
| | GTP cyclohydrolase-2 | | | | | |
| | DARP reductase | | | | | |
| | Bifunctional protein RIB2 | | | | | |
| | DRAP deaminase | | | | | |
| | HTP reductase | | | | | |
| | Lumazine synthase | | | | | |
| | Riboflavin synthase | | | | | |
| | Riboflavin kinase | | | | | |
| | FAD synthase | | | | | |
| Pyridoxal phosphate (Vitamin B6) | D-erythrose-4-phosphate dehydrogenase | | | | | |
| | Erythronate-4-phosphate dehydrogenase | | | | | |
| | Phosphoserine aminotransferase | | | | | |
| | 4-hydroxythreonine-4-phosphate dehydrogenase | | | | | |
| | Pyridoxine 5'-phosphate synthase | | | | | |
| | Pyridoxal kinase | | # | | | |
| | PLP phosphatase | | | | | |
| | Pyridoxine-5'-phosphate oxidase | | | | | |

| Pathway | Enzyme | Insects Gm Ag / Dm Tc | Endosymbionts Wg / Wp Sg | Free-living bacteria Ec / Bs | Plants At / Sb | Yeast/Man Sc / Hs |
|---|---|---|---|---|---|---|
| Nicotinamide adenine dinucleotide (Vitamin B3) | L-aspartate oxidase | | | | | |
| | Quinolinate synthase A | | | | | |
| | Nicotinamide phosphoribosyltransferase | | | | | |
| | Nicotinamidase | | # | | | |
| | Nicotinate phosphoribosyltransferase | | * | | | |
| | Nicotinamide mononucleotide adenylyltransferase | | | | | |
| | NAD synthetase | | | | | |
| | NAD kinase | | | | | |
| | Quinolinate phosphoribosyltransferase | | | | | |
| Co-enzyme A (Vitamin B5) | Branched chain amino acid aminotransferase | | * | | | |
| | Ketopantoate hydroxymethyltransferase | | | | | |
| | Ketopantoate reductase | | | | | |
| | Dihydropyrimidine dehydrogenase | | | | | |
| | Dihydropyrimidinase | | | | | |
| | Beta-ureidopropionase | | | | | |
| | Aspartate 1-decarboxylase | | | | | |
| | Pantothenate synthetase | | | | | |
| | Pantothenate kinase | | | | | |
| | Phosphopantothenate-cysteine ligase | | | | | |
| | Phosphopantothenoylcysteine decarboxylase | | | | | |
| | Phosphopantetheine adenylyltransferase | | | | | |
| | Dephospho-CoA kinase | | | | | |
| Tetrahydrofolate (Vitamin B9) | GTP cyclohydrolase 1 | | | | | |
| | Alkaline phosphatase | | # | | | |
| | Dihydroneopterin triphosphate pyrophosphatase | | | | | |
| | Dihydroneopterin aldolase | | | | | |
| | 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase | | | | | |
| | Para-aminobenzoic acid synthase | | | | | |
| | Aminodeoxychorismate lyase | | | | | |
| | Dihydropteroate synthase | | | | | |
| | Folylpolyglutamate synthase | | | | | |
| | Dihydrofolate reductase | | | | | |
| Biotin (Vitamin B7) | 8-amino-7-oxononanoate synthase | | | | | |
| | DAPA aminotransferase | | | | | |
| | Dethiobiotin synthetase | | | | | |
| | Biotin synthase | | | | | |
| | Biotin protein ligase | | | | | |

**Figure 18. Comparison of B-vitamins and cofactors biosynthesis enzymes in insects, free-living bacteria, endosymbionts, plants, yeast and humans.** Coloured triangles indicate the presence of the respective enzyme in an organism and its absence when white (not coloured). Enzymes encoded by *Glossina* but missing in all the endosymbionts are marked with # while those missing in *Wigglesworthia* are marked with *. Abbreviations: Dm, *Drosophila melanogaster*; Ag, *Anopheles gambiae*; Gm, *Glossina morsitans*; Tc, *Tribolium castenum*; Wg, *Wigglesworthia glossinidia*; Sg, *Sodalis glossinidius*; Wp, *Wolbachia pipientis*; Ec, *Escherichia coli*; Bs, *Bacillus subtilis*; At, *Arabidopsis thaliana*; Sb, *Sorghum bicolor*; Hs, *Homo sapien*; Sc, *Saccharomyces cerevisiae*.

Insects generally have incomplete B-vitamins biosynthesis pathways and most of the enzymes are dedicated to cofactors biosynthesis. In contrast, bacteria, plants and yeast have most enzymes for biosynthesis of B-vitamins and cofactors. For example bacteria, plants and yeast have all the enzymes for riboflavin biosynthesis, while insects have none; all have enzymes for biosynthesis of FAD. *Glossina* has similar enzymes as *D. melanogaster* but differs from *An. gambiae* and *T. castanum* which lack riboflavin kinase and dihydrofolate reductase respectively. B-vitamins and cofactors biosynthesis pathways in insects are therefore generally similar.

On comparison to its endosymbionts and using the free living bacteria as reference, *Glossina* has a number of unique enzymes that the endosymbionts do not encode namely thiamine pyrophosphokinase, pyridoxal kinase, nicotinamidase, and alkaline phosphatase (Figure 18). In addition, when *Glossina* is specifically compared to its obligate endosymbiont *Wigglesworthia*, this bacterium does not encode nicotinate phosphoribosyltransferase and the branched-chain amino acid aminotransferase in NAD and CoA pathways respectively (Figure 18).

In terms of numbers 56 enzymes from *E. coli* genome (Blattner, 1997) were taken as the minimum required for B-vitamins and cofactors biosynthesis. *Glossina* enzymes therefore make up 46.4% (26) while the other insects *Drosophila*, *Anopheles* and *Tribolium* have 44.6% (25), 42.9% (24) and 42.9% (24) enzymes respectively thus *Glossina* has more enzymes than *Anopheles*, *Drosophila* and *Tribolium*. The endosymbionts, *Wiggleswothia*, *Sodalis* and *Wolbachia* have 80.3% (45), 78.6% (44) and 30.4% (17) enzymes respectively. *Wolbachia* has an extremely reduced coding capacity for B-vitamins and cofactors biosynthesis enzymes as compared to the other endosymbionts, an indication of its parasitic

lifestyle. The endosymbionts generally lack some biosynthesis enzymes, an indication that though they have most enzymes, some of their pathways may be incomplete.

Collectively, tsetse has incomplete B-vitamins biosynthesis pathways like other insects and humans whereas plants, free living bacteria and yeast have complete pathways. In contrast, most organisms have complete cofactors biosynthesis pathways. Together, tsetse has the potential to encode enzymes in B-vitamins biosynthesis pathway which the endosymbionts do not encode, an indication of tsetse-endosymbiont cooperation in vitamins biosynthesis through metabolite sharing.

### 4.7 Vitamin transport proteins

Since the B-vitamins biosynthesis pathways were incomplete in tsetse and endosymbionts, and because of experimental indication of vitamin biosynthesis by endosymbionts and import by tsetse (Pais *et al.*, 2008), potential vitamin transporters that could be utilized were sought in tsetse and endosymbionts. For *Glossina*, vitamin transporters were identified in its recently published genome (IGGI, 2014) while in endosymbionts, a search for transport proteins was done in their protein annotations databases namely http://www.ncbi.nlm.nih.gov/genome/proteins/1066?project_id=88075 for *Wigglesworthia*, http://www.ncbi.nlm.nih.gov/genome/proteins/531?project_id=58553 for *Sodalis* and http://www.ncbi.nlm.nih.gov/genome/proteins/11990?project_id=81759 for *Wolbachia*. Further, a homology search similar to the one used to find *Glossina* genes (see section 3.4), was done on the endosymbionts genomes using vitamin transport orthologs from other prokaryotes, then the domains of the resulting hits were compared with the domains of the orthologs used for the search.

Among the endosymbionts, only *Sodalis* was found to have vitamin transport proteins (Figure 19). Some of these proteins have earlier been reported for transport of thiamine (Snyder *et al.*, 2010). In addition, other proteins including a hypothetical protein (Figure 19, panel A) were obtained in *Sodalis* whose domains resemble those of vitamin transport proteins in other bacteria. These proteins have domains that resemble biotin and niacin transporters but no transport protein for the other B-vitamins were identified.

In *Glossina* five vitamin transport proteins have been identified (IGGI, 2014). These are multivitamin transporters with two specific for transport of thiamine and folate as indicated by their domain architecture (Figure 19, panel B).

55

**Figure 19. Domain architecture of *Sodalis* and *Glossina* vitamin metabolites transport proteins.** The domain architectures were obtained using a combination of pfam, prosite and interpro. In panel A, *Sodalis* transport proteins are similar to experimentally validated vitamins transport proteins from other bacteria. The *Sodalis* proteins marked with * are general transport proteins while that marked with ^ is a hypothetical protein. In panel B, the domain organization of putative *Glossina* transport proteins is also similar to experimentally validated proteins from other eukaryotes.

In summary, tsetse encodes six and 20 enzymes for biosynthesis of B-vitamins and cofactors respectively and these have similar domain architecture to other insect orthologs. B-vitamins and cofactors biosynthesis pathways are incomplete in both tsetse and endosymbionts but together, they show complete or near complete pathways. These together with the availability of vitamins transporters in both tsetse and endosymbionts suggests a possible interaction. If accurate, then expression of *Glossina* enzymes implicated in these pathways could vary based on the presence or absence of the endosymbionts, and this was investigated.

## 4.8 Differential expression of *Glossina* B-vitamins and cofactors biosynthesis enzymes

If there is an association between tsetse and endosymbionts in B-vitamins biosynthesis, then there could be potential changes in expression levels of respective genes in tsetse depending on availability of endosymbionts. Therefore differential expression of *Glossina* genes encoding B-vitamins and cofactors biosynthesis enzymes was determinded in the presence and absence of its obligate endosymbiont *Wigglesworthia*, using RNA seq data. Four categories were applied. First, for B-vitamins biosynthesis enzymes specifically absent in *Wigglesworthia*, second, for enzymes that *Glossina* utilizes to generate cofactors from the vitamins, third, for enzymes that are present in *Glossina* and not in the endosymbionts and finally, for vitamin transport proteins in *Glossina* (Figure 20).

B-vitamins biosynthesis enzymes namely GTP cyclohydrolase I, branched chain aminoacid aminotransferase, dihydropyrimidine dehydrogenase and dihydropyrimidinase, missing in *Wigglesworthia* were generally downregulated 1.0-2.6 folds apart from beta-ureidopropionase which was upregulated 3.0 folds (Figure 20). Cofactors biosynthesis enzymes were upregulated in the absence of *Wigglesworthia*. For example, riboflavin kinase, FAD synthase, pyridoxal kinase and phosphopantothenoylcysteine decarboxylase were upregulated by approximately two-folds while thiamine pyrophosphokinase, PNP oxidase, pantothenate kinase, bifunctional coenzyme A synthase and dihydrofolate reductase were upregulated between 1.0-1.5 folds. The enzymes that generate NAD and phosphopantothenate-cysteine ligase in CoA biosynthesis pathway were downregulated by approximately 1.0-1.5 folds. Alkaline phosphatase, pyridoxal kinase and thiamine pyrophosphokinase lacking in the endosymbionts – with reference to the free living bacteria – in addition to three vitamin transporter proteins were upregulated approximately 1.0-5.0 fold.

In summary, *Glossina* has incomplete pathways for biosynthesis of B-vitamins, but complete machinery for biosynthesis of cofactors. In addition it encodes B-vitamins biosynthesis enzymes lacking in endosymbionts; these are downregulated in absence of *Wiggleswothia*, likely because of accumulation of metabolites utilized by *Wiggleswothia*. Conversly, most enzymes that generate cofactors from B-vitamins are upregulated, an indication of fly compensation for deficit in some metabolites shared with endosymbionts, hence supporting interaction. The vitamin transport proteins in both tsetse and endosymbionts may facilitate this interaction and are upregulated in the absence of *Wiggleswothia*. Together, the data suggest a possible interaction between tsetse and its endosymbionts in biosynthesis of B-vitamins.
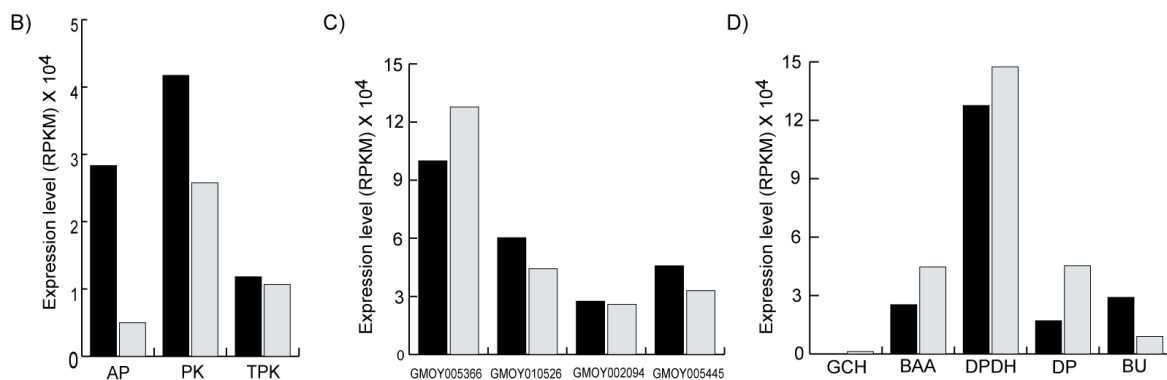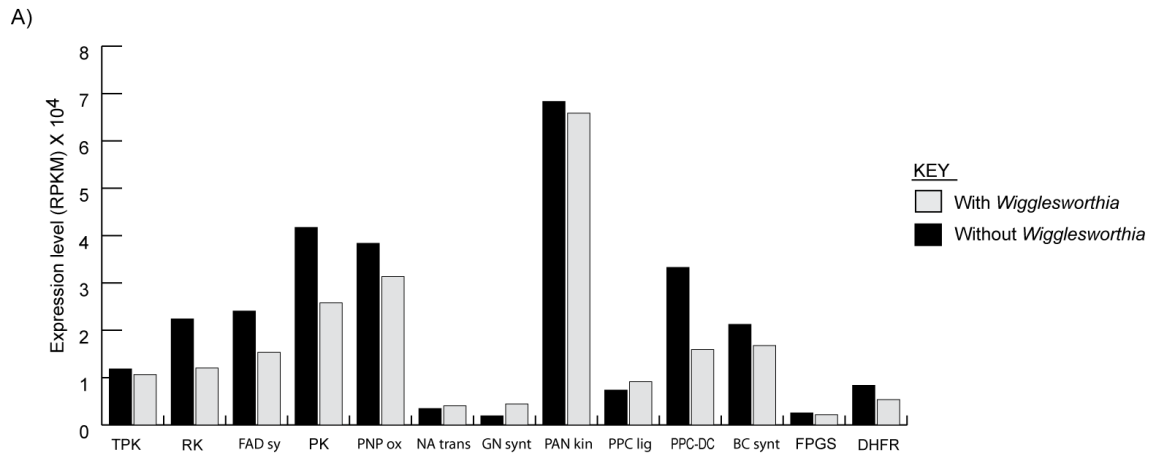
**Figure 20. Differential expression level of enzymes in *Glossina*.** RNA seq data from *Wigglesworthia* free (ampicillin treated) and tsetse with *Wigglesworthia* were used to compare gene expression level of B-vitamins and cofactors biosynthesis genes. Panel A represents enzymes that generate cofactors from vitamins. Panel B are enzymes present in *Glossina* but absent in all the endosymbionts, panel C are vitamin transport proteins in *Glossina* while panel D represents B-vitamins biosynthesis enzymes absent in *Wigglesworthia*. Abbreviations: TPK, Thiamine pyrophosphokinase; RK, riboflavin kinase; FAD synt, flavin adenine dinucleotide synthase; PK, pyridoxal kinase; PNP ox, pyridoxine-5'-phosphate oxidase; NA trans, nicotinamide mononucleotide adenylyltransferase; GN synt, glutamine dependent nicotinamide adenine dinucleotide synthetase; PAN kin, pantothenate kinase; PPC lig, phosphopantothenate-cysteine ligase; PPC-DC, phosphopantothenoylcysteine decarboxylase; BC synt, bifunctional coenzyme A synthase; FPGS, folylpolyglutamate synthase; DHFR, dihydrofolate reductase; AP, alkaline phosphatase; PK, pyridoxal kinase; TPK, thiamine pyrophosphokinase; GCH, GTP cyclohydrolase; BAA, branched chain aminoacid aminotransferase; DPDH, dihydropyrimidine dehydrogenase; DP, dihydropyrimidinase; BU, beta ureidopropionase.

# CHAPTER FIVE

## DISCUSSION

### 5.1 Annotation of *Glossina* genes involved in B-vitamins and cofactors biosynthesis

A total of 26 enzymes involved in B-vitamins (six) and cofactors (20) biosynthesis were annotated from the recently sequenced *Glossina morsitans morsitans* genome (IGGI, 2014). The annotation process involved bioinformatics analyses namely blast searches, domain prediction, protein physicochemical properties determination and RNA seq analysis. These were complemented with PCR amplification, cloning, sequencing and analysis of selected annotated genes. All these analyses ensured robustness of the annotation process.

Firstly, all possible B-vitamins and cofactors biosynthesis pathways were determined using experimentally validated enzymes known to participate in these pathways (Table S1) and global biosynthesis pathways constructed (Figure 14). Computer predicted orthologs that had similar domain organization to the experimentally validated ones (Figure 15) were also included. These pathways provided a blueprint of all the possible systems of vitamin biosynthesis utilized and orthologs for searches, hence the *Glossina* genome exhaustively interrogated.

Secondly, all the orthologs were used to search the genome of *Glossina*. This ensured that the *Glossina* genes were identified multiple times (as well as through reverse blast) hence all possible pathways were covered in the analysis. In addition all the resulting *Glossina* proteins were confirmed to have similar domain organization and physicochemical properties as experimentally validated orthologs eliminating false homologous sequences (Figure 16). Moreover they were not similar to endosymbiont orthologs since the sequenced isolate was devoid of endosymbionts and most endosymbiont enzymes were known thus enabling comparison (Figure 18).

Thirdly, a transcriptome obtained independently from the sequenced tsetse genome was mapped on the genome and the putative annotated B-vitamins and cofactors biosynthesis genes identified. This guaranteed that the annotated genes are expressed hence possibly functional, and the multiple RNA seq data used were devoid of genomic DNA contaminant and gave identical mapping results. In addition, the differential expression mediated by availability of endosymbiont also suggested accurate annotation.

Finally, the successful amplification, cloning and sequencing of selected annotated genes using genomic material from *G. m. morsitans* strain different from the sequenced Yale strain suggests that the sequenced genome is of high quality hence annotations were most likely valid. The targeted amplifications were similar to bioinformatics predicted gene models, again confirming presence and authenticity of the annotated *Glossina* genes.

Together, these gene models were determined using multiple approaches including sequence homology and domain similarity to experimentally validated orthologs, by presence of their corresponding RNA and PCR amplification, indicating presence and authenticity. Moreover, differential expression data suggests that the annotated genes encode for true B-vitamins and cofactors biosynthesis enzymes. Therefore the annotation data most likely represents all or most of the enzymes involved in B-vitamins and cofactors biosynthesis in tsetse.

## 5.2 Insects exhibit incomplete B-vitamin biosynthesis pathways

B-vitamins are precursors of cofactors which are components of most enzymes and are therefore vital for life. Free living organisms have means of synthesizing these nutrients *de novo* and therefore have complete biosynthesis pathways. For example plants, free living bacteria and yeast have complete B-vitamins biosynthesis pathways (Figure 18), making these organisms dietary sources of vitamins.

In contrast, insects, humans and some bacteria (parasitic or symbiotic) have incomplete B-vitamins biosynthesis pathways (Figure 18). They therefore outsource these nutrients from diet or through symbiotic associations. For example, insects like mosquitoes, fruit flies and flour beetle feed on plant materials which are rich in vitamins, while obligate feeders like tsetse flies and bat flies associate with bacteria symbionts which synthesize and provide these B-vitamins (Hosokawa *et al.*, 2012; Wang *et al.*, 2013). Humans also have co-evolved with numerous symbiotic bacteria which are a source of vitamin B and K (Cummings and Macfarlane, 1997).

Tsetse has co-evolved with three endosymbionts particularly the obligate endosymbiont *Wigglesworthia* for about 50-80 million years (Chen *et al.*, 1999). *Wigglesworthia* has retained the capacity to biosynthesize most B-vitamin which it provides to its invertebrate host (Akman *et al.*, 2002; Snyder *et al.*, 2010). Other insects that have lost some metabolic capacity but associate with symbionts to compensate this loss include aphids (retained *Buchnera* to provide essential amino acids) (Shigenobu *et al.*, 2000) and carpenter ants (retained *Blochmannia* to provide nitrogen and sulphur compounds) (Gil *et al.*, 2003). These

co-evolution and association has enabled hosts to survive on restricted diets in their ecological niches since the endosymbionts provide the nutrients that the hosts are unable to biosynthesize or are not available in their restricted diets (Aksoy *et al.*, 2005).

Cofactors are synthesized from the B-vitamins. All organisms analysed have complete cofactors biosynthesis pathways regardless of their lifestyle (Figure 18). This means that provided that an external source of B-vitamins is available, all the organisms can independently synthesize cofactors. For example, *Glossina* cured of its obligate endosymbiont can still be sustained if it is provided with an alternative source of B-vitamins e.g. a B-vitamins supplemented blood meal (Pais *et al.*, 2008). This has enabled study of the effects of individual endosymbionts by performing specific elimination of endosymbionts using antibiotics (Weiss *et al.*, 2006, 2011; Pais *et al.*, 2008; Wang *et al.*, 2013). In these studies, tsetse blood meal is supplemented with a cocktail of B-vitamins and the flies remain alive longer than their aposymbiotic counterparts lacking the supplemented blood meal. This is a clear indication that endosymbionts have a biochemical role in vitamin biosynthesis.

Insects therefore have potential to biosynthesize cofactors if provided with vitamins. The reason why these insects' B-vitamins and cofactors biosynthesis pathways are similar yet they have different sources of B-vitamins can only be speculated. Their ancestors may have outsourced B-vitamins from diet e.g. plant materials, a feature that is retained despite acquisition of different lifestyles. For example, *Drosophila* and *Tribolium* have maintained a plant meal lifestyle while *Glossina* acquired a blood meal lifestyle, but *Anopheles* retained an intermediary lifestyle i.e. feeding on both plant materials and blood. *Glossina* became an obligate blood feeder and therefore compensated for nutrients e.g. vitamins previously obtained from plant materials by acquiring an obligate endosymbiont, *Wigglesworthia*. In addition, to ensure survival of its young, it acquired an adenotropic viviparity reproduction and mammary feeding system that allow direct transfer of endosymbionts to larvae, hence no limit in endosymbiont-dependent nutrient supply. In contrast, *Anopheles*, *Drosophila* and *Tribolium* have no obligate endosymbionts but have other endosymbionts. Specifically, all have *Wolbachia* which manipulates host reproduction to favour its spread among insect populations (Mateos *et al.*, 2006; Hughes *et al.*, 2011; Goodacre *et al.*, 2013) and *Drosophila* in addition hosts *Spiroplasma* which protects against pathogenic infections (Mateos *et al.*, 2006). *Anopheles* hosts *Serratia, Enterobacter, and Asaia* (Minard *et al.*, 2013). *Serratia* and *Enterobacter* hemolytic enzymes suggests a role in blood digestion while *Asaia* may provide acetic acid (Minard *et al.*, 2013).

61

Apart from *Wolbachia* tsetse endosymbionts *Sodalis* and *Wigglesworthia* also have the potential of making their own cofactors (Akman *et al.*, 2002; Toh *et al.*, 2006; Pais *et al.*, 2008; Snyder *et al.*, 2010). Since tsetse B-vitamins and cofactors biosynthesis pathways are still being unravelled, its blood meal is deficient in B-vitamins and its endosymbionts potentially biosynthesize both B-vitamins and cofactors, it could follow that endosymbionts provide tsetse with both B-vitamins and cofactors. This study has therefore clarified that tsetse has complete cofactors biosynthesis pathways and has the potential to biosynthesize its own cofactors. However it remains unknown whether the endosymbionts also provide tsetse with cofactors, but the facts that tsetse has cofactors biosynthesis pathways, and nutritional and endosymbiont knock-out studies (Pais *et al.*, 2008; Weiss *et al.*, 2011) suggest that tsetse biosynthesizes cofactors, makes this option highly likely.

In summary, insects and humans have incomplete B-vitamins biosynthesis pathways. This is in contrast with free living organisms like plants, yeast and some bacteria. Organisms that lack biosynthesis capabilities either acquire vitamins from diet and/or from symbiotic associations that permit supply or co-synthesis hence incomplete individual biosynthetic pathways. On association and hence integration of the pathways, the resulting machinery has capabilities of sharing precursors and metabolites involved in biosynthesis, hence a biochemical interaction responsible for the intimate association.

## 5.3 Possible association between *Glossina* and its endosymbionts in B-vitamins biosynthesis

The intimate association of tsetse and endosymbionts has been demonstrated experimentally (Pais *et al.*, 2008; Snyder *et al.*, 2010; Weiss *et al.*, 2011; Wang *et al.*, 2013). This association at least involves vitamins provision and/or metabolism by endosymbionts (Akman *et al.*, 2002; Pais *et al.*, 2008; Snyder *et al.*, 2010; Rita *et al.*, 2012) while tsetse provides metabolic precursors in addition to housing its endosymbionts. The bioinformatics analysis of B-vitamins biosynthesis enzymes in tsetse and endosymbionts (Figures 11, 12, 18, and A1), indicate potential interaction at molecular and biochemical levels due to the following reasons.

First, combination of tsetse and endosymbionts vitamin biosynthesis pathways provides complete or near complete pathways, justifying the role of endosymbiont in tsetse immunity, longevity, fecundity and digestion (Pais *et al.*, 2008; Weiss *et al.*, 2011). Here there is likelihood that metabolic intermediates from both partners are utilized to synthesize final

products, specifically vitamins that are utilized individually for cofactor biosynthesis and subsequently in enzymatic activity. For example *Glossina* contributes in the first steps of pantothenate biosynthesis pathway in *Wigglesworthia* and folate biosynthesis pathway in both *Wigglesworthia* and *Sodalis* (Figures 21 and A1). As a result, in pantothenate biosynthesis pathway, *β*-alanine and *α*-ketovalerate are possibly synthesized by the fly then passed on to *Wigglesworthia* pantothenate pathway. Similarly, in folate biosynthesis pathway, dihydroneopterin is possibly made in the fly then imported by both *Wigglesworthia* and *Sodalis* for the synthesis of folate.

Apart from folate biosynthesis pathway, all the incomplete B-vitamins pathways in *Wigglesworthia* could also be complemented by *Sodalis*, but since *Sodalis* is commensal, it is likely that tsetse and not *Sodalis* complements *Wigglesworthia* pathways. Further, *Wigglesworthia* has intimate association with tsetse than it has with *Sodalis* because it is the only symbiont localized in the bacteriocytes. However the involvement of *Sodalis* may also be a possibility.

Second, variations in tsetse gene expression of B-vitamins and cofactors biosynthesis enzymes in presence and absence of *Wigglesworthia* (Figure 20) suggest an interaction. Here, there is a possibility of feedback mechanism. In absence of endosymbiont, expression levels of tsetse B-vitamins biosynthesis enzymes involved in early steps of biosynthesis before transfer of metabolites to endosymbionts are downregulated, possibly due to negative inhibition by increased products. In the case where there is upregulation (e.g. *β*-ureidopropionase), there is positive feedback due to reduced substrate level. In contrast, expression of cofactors biosynthesis enzymes increase in response to low metabolite levels available. For few cases where downregulation was observed, negative feedback due to lack of metabolites could be responsible, analogous to *E. coli lac* operon where genes are downregulated in the absence of lactose (Busby and Ebright, 1999). Though this is speculative, it could be addressed by comparing these expression levels to those of tsetse initially devoid of endosymbionts but are on vitamin-supplemented blood meal diet. In addition, the role of the other endosymbionts namely *Sodalis* and *Wolbachia* in these pathways can not be overruled and remains unknown.

Third, tsetse have putative vitamin transporters (IGGI, 2014) that are possibly used for uptake of vitamins and vitamin-like metabolic intermediates and are generally upregulated in the absence of *Wigglesworthia* (Figure 20). Tsetse cured of *Wigglesworthia* incurs fitness cost in

reproduction, immunity, digestion and growth and this cost is partially alleviated by supplementation of the tsetse blood meal with a cocktail of B-vitamins, an indication of *Wigglesworthia* role in providing B-vitamins (Pais *et al.*, 2008). These transporters could be implicated in vitamin and metabolites transport in the fly's cells and hence available for synthesis of cofactors. Through this pathway, intracellular endosymbionts also access vitamins hence transporters in *Sodalis* especially those for thiamine import, given that this endosymbiont lacks a *de novo* thiamine biosynthesis pathway (Toh *et al.*, 2006; Snyder *et al.*, 2010). Upregulation of transporter expression could be due to scavenging for limited vitamins and/or biosynthesis intermediates to meet the insect's needs. In complete absence of the vitamins, fitness in compromised.

In summary, tsetse has incomplete B-vitamins biosynthesis pathways and complete cofactors biosynthesis pathways similar to other insects, and most likely outsource for B-vitamins. To compensate for its restricted blood meal deficient in B-vitamins, tsetse lives in symbiosis with the obligate endosymbiont *Wigglesworthia*, an association that permits collaboration in vitamins biosynthesis. This is supported by complete or near complete pathways on integration of tsetse and endosymbiont biosynthetic pathways hence possible interaction. Further, gene expression of enzymes and transporters vary with endosymbiont availability, supporting interaction.

### 5.4 Hypothetical model for tsetse-endosymbiont interaction

If tsetse and endosymbionts share in vitamin biosynthesis, what mechanism could be employed? Since most endosymbionts are intracellular, their main source of nutrients and metabolic precursors most likely is the host, which gets its requirements from blood meal. Consequently, the endosymbionts import some metabolites as precursors for vitamin biosynthesis (Figure 21) using substrate specific transporters (Figure 12) (Akman *et al.*, 2002; Toh *et al.*, 2006). Subsequently the metabolites are utilized in B-vitamins biosynthesis and part of the end products i.e. vitamins, are exported to the fly for utilization in cofactors biosynthesis (Figure 21).

With the fly's potential to catalyze the first steps of pantothenate and folate biosynthesis lacking in endosymbionts, it has the capacity to prime synthesis of these vitamins. This suggests that the fly synthesizes *β*-alanine, *α*-ketoisovalerate and dihydroneopterin that are subsequently exported to the endosymbionts through transporters (Figure 21). Once in endosymbionts, these metabolites are utilized in biosynthesis of pantothenate and folate, part

64

of which are exported to the insect cell. These vitamins are utilized in cofactor biosynthesis by both partners, thus a possible biochemical interaction. In the absence of *Wigglesworthia*, these metabolites cannot be processed further due to a disrupted integrated pathway, hence no end products (vitamins) for cofactors biosynthesis, resulting into poor digestion and reproduction and subsequently death of tsetse.
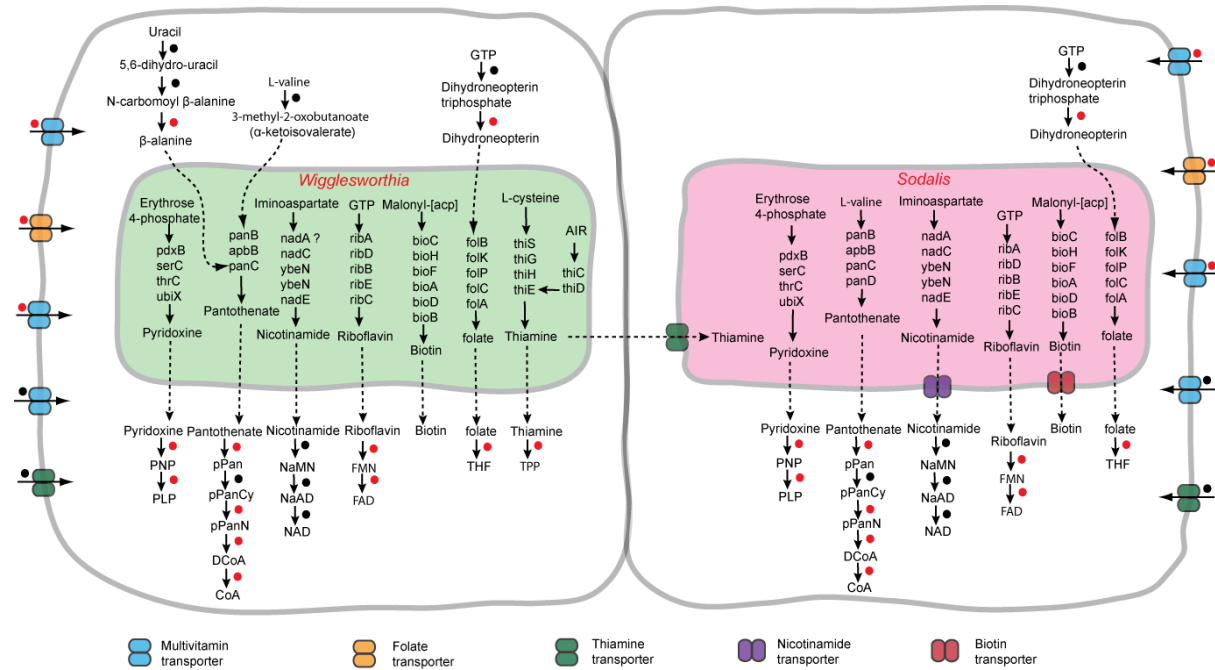


**Figure 21. An illustration of the interaction between a *Glossina* cells and cytoplasmic endosymbionts in biosynthesis of B-vitamins and cofactors.** The pathways presented were obtained after integrating *Glossina*, *Wigglesworthia* and *Sodalis* pathways. Dotted and block arrows represent transport and enzyme reactions respectively while red and black dots indicate upregulation and downregulation of protein respectively. Transport proteins are accompanied with arrows to show direction of movement of metabolites.

Transport proteins in *Wigglesworthia* would facilitate export of vitamins into the tsetse cytosol but orthologs have not been identified and may have possibly been lost (Akman *et al.*, 2002; Rita *et al.*, 2012) due to this association, or an unknown machinery could be in operation. Other symbionts e.g. the obligate endosymbiont of aphids *Buchnera aphidicola* and *Blochmannia floridanus* of carpenter ants have also lost B-vitamin transporters (Gil *et al.*, 2003; van Ham *et al.*, 2003; Price *et al.*, 2011). *Sodalis* however has retained transporters for thiamine (Snyder et al., 2010) possibly because it lacks a complete thiamine biosynthesis pathway and has to acquire it from tsetse and other endosymbionts (Figure 21). In addition, it encodes other putative transporters for biotin and niacin (Figures 19 and 21). *B. aphidicola*

and *B. floridanus* have retained amino acid transporters to facilitate export of essential amino acids which the hosts can not synthesize and in return, they import non-essential ones which symbionts can not synthesize (Gil *et al.*, 2003; van Ham *et al.*, 2003; Price *et al.*, 2011). These suggest the possibility of movement of metabolites across the host and endosymbionts, a scenario that could be replicated in tsetse-endosymbionts association hence the presence of some transporters in intracellular endosymbionts.

In the absence of endosymbionts, B-vitamins can be supplied to tsetse through supplemented blood meal. In this case, tsetse cells absorb B-vitamins from the extracellular matrix using multivitamin, thiamine and folate transporters (Figure 21). These transporters are generally upregulated during vitamin deficiency (absence of *Wigglesworthia*) to increase intake from extracellular sources. Since other B-vitamin transporters have not been identified, it is possible that other means are used to absorb the remaining B-vitamins not represented by respective transporters. Similar to the B-vitamins supplied by the endosymbionts, the supplemented forms are also metabolized into cofactors and incorporated into enzyme systems.

Based on pathway analysis of tsetse and its endosymbionts, differential expression of tsetse enzymes at different points of the pathways and regulation of its transport proteins in the absence of its obligate endosymbiont, it can be hypothesized that endosymbionts and tsetse associate in vitamin biosynthesis using a machinery that relies on intracellular metabolites, vitamin transport and integrated biosynthesis pathway hence biochemical interaction.

## 5.5 Potential utilization of tsetse-endosymbiont biomolecular interaction

Endosymbionts can be exploited as targets for control of trypanosomiasis (Aksoy, 2000). For example, paratransgenesis involves use of transgenic endosymbionts to either kill the fly or prevent the fly from transmitting parasites (Aksoy *et al.*, 2008). In tsetse flies, the endosymbiont *Sodalis*, has been preferred for paratransgenesis because of its ability to culture and transform *in vitro* and its versatility in the insect body. In this method, transgenic *Sodalis* is propagated into tsetse population where it expresses a trypanocidal agent that target procyclic trypanosomes, preventing their establishment, hence blocking transmission (Medlock *et al.*, 2013). Though promising, this method has not been developed to success owing to challenges of maintaining the flies bearing transformed *Sodalis*. Therefore, more insight on tsetse-endosymbionts interaction is necessary hence our study.

If the hypothesized interaction is experimentally validated, then there are various ways that it can be exploited for improvement of paratransgenesis as a control strategy. Two approaches can be considered. First, the mechanism of endosymbiont export of vitamin intermediates coupled to tsetse exocytic machinery, can be utilized to export endosymbiont (in this case *Sodalis*)-derived trypanocidal agents into the various tsetse structures (such as mid gut and salivary glands) where trypanosomes develop, killing the parasites hence blocking transmission. Here, transgenic endosymbionts e.g. the amenable *Sodalis* is genetically modified to express and export trypanocidal agents into tsetse cells which subsequently export these agents into the fly structures where they act against trypanosomes. This will prevent parasite establishment in the fly and block transmission. Alternatively, since trypanosomes do not biosynthesize B-vitamins *de novo* (Berriman *et al.*, 2005) and most likely acquire vitamins from tsetse and endosymbionts, blocking this supply could be a potential control strategy. This approach however may result in death of the fly, hence affecting ecological balance. This limitation is not observed in the former approach. These however will come clear after thorough understanding of tsetse-endosymbiont-trypanosome interactions, which can be fast tracked due to available genomes of the three players.

From this study, the annotation suggests that at least tsetse has B-vitamins and cofactors biosynthesis pathways. B-vitamins pathways are however not complete in tsetse, a phenomenon common to insects and humans too. On the other hand, free living organisms like plants, yeast and *E.coli* have complete pathways and therefore synthesize B-vitamins *de novo*. Tsetse which does not feed on vitamin-rich plant material like mosquitoes and fruit flies, benefits from its association with bacterial endosymbionts to acquire B-vitamins deficient in its blood meal. This symbiosis most likely involves B-vitamins biosynthesis, where tsetse provides biosynthesis precursors to endosymbionts for vitamins biosynthesis, in return for sharing the synthesized vitamins; a strong indication of biochemical interaction. All organisms however have complete cofactors biosynthesis pathways and should synthesize their own cofactors if provided with a vitamin source. This metabolic symbiosis in biosynthesis of B-vitamins is supported by pathway complementarity, presence of vitamins transport proteins and endosymbiont-dependent expression of B-vitamins and cofactors biosynthesis enzymes in tsetse. Nutritional symbiosis is also common among insects like aphids, which feed on amino acids-deficient plant materials, whose endosymbionts provide amino acids. With further studies, this interaction at biochemical and molecular level can be targeted in paratransgenesis to prevent parasite transmission and/or vector control.

# CHAPTER SIX

## CONCLUSIONS AND RECOMMENDATIONS

### 6.1 Conclusions

The *G. m. morsitans* genome has 34 genes which encode for 26 B-vitamins and cofactors biosynthesis enzymes; eight genes encode isoenzymes. Cofactors biosynthesis pathways are complete in tsetse as observed in other groups of organisms including insects (*Drosophila*, *Anopheles* and *Tribolium*), plants and yeast. However, B-vitamins biosynthesis pathways are incomplete, a phenomenon also observed in insects, humans and endosymbiotic bacteria. One aspect of association between tsetse and its endosymbionts has been shown to be on vitamin metabolism, and comparative analysis suggests a likelihood of an association in biosynthesis of B-vitamins. This is supported by first, complete or near complete pathways only when tsetse and endosymbionts systems are integrated; second, presence of vitamin transporters in tsetse and intracellular endosymbionts; and finally, differential gene expression dependent on availability of endosymbionts. This study therefore provides insight into possible metabolic interdependence between tsetse and its endosymbionts and provides possible avenues for control of trypanosomiasis, for example by exploiting the transport mechanism of these B-vitamins. Alternatively, the potential that trypanosomes obtain B-vitamins from the tsetse-endosymbiont symbiotic association enables a possible strategy that blocks trypanosomes' access to B-vitamins and thus inhibit development of procyclic trypanosomes hence prevent transmission of trypanosomes by tsetse. However this is speculative and would require further validation to have an accurate picture of the association.

### 6.2 Recommendations

The observations made in this study suggest potential endosymbiont-tsetse interaction in B-vitamin metabolism. This was carried out using computational approach and limited wet lab experiments. Going forward, improvement of our knowledge in this area is vital for development of paratransgenesis and/or development of novel control strategies. To exploit this potential, the following are recommended. First, confirming experimentally if the annotated genes are truly involved in vitamin metabolism. This can be through functional analysis by applying RNA interference (RNAi) of the putative genes, Western blotting for enzymes and enzyme kinetics studies among others. Second, the association in vitamin metabolism must be validated. The tools applicable include tracing labelled vitamins biosynthesis precursors in the two systems, combination of RNAi and provision of vitamin precursor supplemented blood meal and gene expression studies using quantitative PCR.

Third is to determine the biochemical and molecular interaction of the other endosymbionts with tsetse and trypanosomes. Here, metabolic and immunological processes, and vector competence among other factors could be considered. Finally, understanding tsetse-endosymbionts-trypanosomes interaction would provide more insight in identifying target molecules for use in paratransgenesis. This will be accelerated by the availability of the genomes of the three players and improvement in functional genomics.

**REFERENCES**

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, and Aksoy S. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature Genetics*, **32**: 402–407.

Aksoy S. (1995). *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *International Journal of Systematic Bacteriology*, **45**: 848–851.

Aksoy S. (2000). A Haven for Microorganisms. *Parasitology Today*, **16**: 125–126.

Aksoy S. (2003). Control of tsetse flies and trypanosomes using molecular genetics. *Veterinary Parasitology*, **115**: 125–145.

Aksoy S, Berriman M, Hall N, Hattori M, Hide W, and Lahane M. (2005). A case for a *Glossina* genome project. *Trends in Parasitology*, **21**: 107–111.

Aksoy S, Chen X, and Hypsa V. (1997). Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera:*Glossinidae*). *Insect Molecular Biology*, **6**: 183–90.

Aksoy S, and Rio RVM. (2005). Interactions among multiple genomes: tsetse, its symbionts and trypanosomes. *Insect Biochemistry and Molecular Biology*, **35**: 691–698.

Aksoy S, Weiss B, and Attardo G (2008). Paratransgenesis applied for control of tsetse transmitted sleeping sickness. *Advances in Experimental Medicine and Biology*, **627**: 35–48.

Allsopp R, and Hursey BH. (2004). Insecticidal control of tsetse. In P. H. Maudlin, Holmes, and M. A. Miles (Eds.), *The Trypanosomiases* (pp. 491–507). CABI Publishing.

Anene B, Onah D, and Nawa Y. (2001). Drug resistance in pathogenic African trypanosomes: what hopes for the future? *Veterinary Parasitology*, **96**: 83–100.

Attardo G, Lohs C, Heddi A, Alam UH, Yildirim S and Aksoy S. (2008). Analysis of milk gland structure and function in *Glossina morsitans*: milk protein production, symbiont populations and fecundity. *Journal of Insect Physiology*, **54**: 1236–1242.

Attardo GM, Benoit JB, Michalkova V, Yang G, Roller L, Bohova J, Takáč P and Aksoy S. (2012). Analysis of lipolysis underlying lactation in the tsetse fly, *Glossina morsitans*. *Insect Biochemistry and Molecular Biology*, **42**: 360–370.

Attardo GM, Guz N, Strickler-Dinglasan P and Aksoy S. (2006). Molecular aspects of viviparous reproductive biology of the tsetse fly (*Glossina morsitans morsitans*): regulation of yolk and milk gland protein synthesis. *Journal of Insect Physiology*, **52**: 1128–1136.

Attardo GM, Strickler-Dinglasan P, Perkin S, Caler E, Bonaldo MF, Soares MB, El-Sayeed N and Aksoy S. (2006). Analysis of fat body transcriptome from the adult tsetse fly, *Glossina morsitans morsitans*. *Insect Molecular Biology*, **15:** 411–424.

Balasegaram M, Young H, Chappuis F, Priotto G, Raguenaud ME, and Checchi F. (2009). Effectiveness of melarsoprol and eflornithine as first-line regimens for gambiense sleeping sickness in nine Médecins Sans Frontières programmes. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **103**: 280–290.

Beard CB, O'Neill SL, Mason P, Mandelco L, Woese CR, Tesh RB, Richards FF and Aksoy S. (1993). Genetic transformation and phylogeny of bacterial symbionts from tsetse. *Insect Molecular Biology*, **1**:123–131.

Begley TP, Chatterjee A, Hanes JW, Hazra A and Ealick SE. (2008). Cofactor biosynthesis - still yielding fascinating new biological chemistry. *PMC*, **12**: 118–125.

Begley TP, Kinsland C and Strauss E. (2001). The biosynthesis of coenzyme A in bacteria. *Vitamins and Hormones*, **61**: 157–171.

Berrang L. (2007). Civil conflict and sleeping sickness in Africa in general and Uganda in particular. *Conflict and Health*, **1**: 6.

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard, E. Caler, N. E. Hamlin, B. Haas, U. Böhme, L. Hannick, M. A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U. C. M. Alsmark, C. Arrowsmith, R. J. Atkin, A. J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T.-J. Chillingworth, C. Churcher, L. N. Clark, C. H. Corton, A. Cronin, R. M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M. C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B. R. Harris, H. Hauser, J. Hostetler, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A. X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. Macleod, P. J. Mooney, S. Moule, D. M. A. Martin, G. W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C. S. Peacock, J. Peterson, M. A. Quail, E. Rabbinowitsch, M.-A. Rajandream, C. Reitter, S. L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A. J. Simpson, L. Tallon, C. M. R. Turner, A. Tait, A. R. Tivey, S. Van Aken, D. Walker, D. Wanless, S. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M. D. Adams, T. M. Embley, K. Gull, E. Ullu, J. D. Barry, A. H. Fairlamb, F. Opperdoes, B. G. Barrell, J. E. Donelson, N. Hall, C. M. Fraser, S. E. Melville, and El-Sayed, N. M. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science*, **309**: 416–422.

Blattner FR. (1997). The Complete Genome Sequence of *Escherichia coli* K-12. *Science*, **277**: 1453–1462.

Brightwell R, Dransfield RD and Kyorku C. (1991). Development of a low-cost tsetse trap and odour baits for *Glossina pallidipes* and *G. longipennis* in Kenya. *Medical and Veterinary Entomology*, **5**: 153–164.

Busby S and Ebright RH. (1999). Transcription activation by catabolite activator protein (CAP). *Journal of Molecular Biology*, **293**: 199–213.

Calvitti M, Moretti R, Skidmore A R and Dobson SL. (2012). *Wolbachia* strain wPip yields a pattern of cytoplasmic incompatibility enhancing a *Wolbachia*-based suppression strategy against the disease vector *Aedes albopictus*. *Parasites and Vectors*, **5**: 254–264.

Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell BG, Rajandream MA. (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics (Oxford, England)*, **24**:, 2672–2676.

Chen X, Li S and Aksoy S. (1999). Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *Journal of Molecular Evolution*, **48**:49–58.

Cheng Q and Aksoy S. (1999). Tissue tropism, transmission and expression of foreign genes in vivo in midgut symbionts of tsetse flies. *Insect Molecular Biology*, **8**: 125–132.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M and Robles M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**: 3674–3676.

Cummings JH and Macfarlane GT. (1997). Role of intestinal bacteria in nutrient metabolism. *Clinical Nutrition*, **16**: 3–11.

Dale C and Maudlin I. (1999). *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. *International Journal of Systematic Bacteriology*, **49**: 267–275.

Doudoumis V, Tsiamis G, Wamwiri F, Brelsfoard C, Alam U, Aksoy E, Dalaperas S, Abd-Alla A, Ouma J, Takac P, Aksoy S, Bourtzis K. (2012). Detection and characterization of *Wolbachia* infections in laboratory and natural populations of different species of tsetse flies (genus *Glossina*). *BMC Microbiology*, **12**: S3.

Douglas AE. (1989). Mycetocyte symbiosis in insects. *Biological Reviews of the Cambridge Philosophical Society*, **64**: 409–434.

Dransfield RD, Williams BG and Brightwell R. (1991). Control of tsetse flies and trypanosomiasis: Myth or reality? *Parasitology Today*, **7**: 287–291.

Edwards MA, Kaufman ML and Storvick CA. (1957). Microbiologic assay for the thiamine content of blood of various species of animals and man. *The American Journal of Clinical Nutrition*, **5**: 51–55.

FAO. (2013). Animal production and health. Retrieved August 12, 2013, from Food and Agriculture Organization of the United Nations: http://www.fao.org/ag/againfo/programmes/en/paat/disease.html

Fèvre EM, Wissmann BV, Welburn SC and Lutumba P. (2008). The burden of human African trypanosomiasis. *PLoS Neglected Tropical Diseases*, **2**: e333.

Fischer M, Haase I, Feicht R, Richter G, Gerhardt S, Changeux J, Bacher A. (2002). Biosynthesis of riboflavin, **526**: 519–526.

Gabelli SB, Bianchet M, Xu W, Dunn CA, Niu ZD, Amzel LM and Bessman MJ. (2007). Structure and function of the *E. coli* dihydroneopterin triphosphate pyrophosphatase: a Nudix enzyme involved in folate biosynthesis. *Structure*, **15**: 1014–1022.

Gil R, Silva FJ, Zientz E, Delmotte F, González-Candelas F, Latorre A and Moya A. (2003). The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**: 9388–9393.

Goodacre SL, Fricke C and Martin OY. (2013). A screen for bacterial endosymbionts in the model organisms *Tribolium castaneum*, *T. confusum*, *Callosobruchus maculatus*, and related species. *Insect Science*. doi:10.1111/1744-7917.12096

Hanson AD and Gregory JF. (2011). Folate biosynthesis, turnover, and transport in plants. *Annual Review of Plant Biology*, **62**: 105–125.

Hassanali A, Herren H, Khan ZR, Pickett J and Woodcock CM. (2008). Integrated pest management: the push-pull approach for controlling insect pests and weeds of cereals, and its potential for other agricultural systems including animal husbandry. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **363**: 611–621.

Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A and Werren JH. (2008). How many species are infected with *Wolbachia*?-A statistical analysis of current data. *FEMS Microbiology Letters*, **281**: 215–220.

Holmes P. (2013). Tsetse-transmitted trypanosomes - Their biology, disease impact and control. *Journal of Invertebrate Pathology*. doi:10.1016/j.jip.2012.07.014.

Holt R, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR P. Wincker, A. G. Clark, J. M. C. Ribeiro, R. Wides, S. L. Salzberg, B. Loftus, M. Yandell, W. H.

Majoros, D. B. Rusch, Z. Lai, C. L. Kraft, J. F. Abril, V. Anthouard, P. Arensburger, P. W. Atkinson, H. Baden, V. de Berardinis, D. Baldwin, V. Benes, J. Biedler, C. Blass, R. Bolanos, D. Boscus, M. Barnstead, S. Cai, A. Center, K. Chaturverdi, G. K. Christophides, M. a Chrystal, M. Clamp, A. Cravchik, V. Curwen, A. Dana, A. Delcher, I. Dew, C. a Evans, M. Flanigan, A. Grundschober-Freimoser, L. Friedli, Z. Gu, P. Guan, R. Guigo, M. E. Hillenmeyer, S. L. Hladun, J. R. Hogan, Y. S. Hong, J. Hoover, O. Jaillon, Z. Ke, C. Kodira, E. Kokoza, A. Koutsos, I. Letunic, A. Levitsky, Y. Liang, J.-J. Lin, N. F. Lobo, J. R. Lopez, J. a Malek, T. C. McIntosh, S. Meister, J. Miller, C. Mobarry, E. Mongin, S. D. Murphy, D. a O'Brochta, C. Pfannkoch, R. Qi, M. a Regier, K. Remington, H. Shao, M. V Sharakhova, C. D. Sitter, J. Shetty, T. J. Smith, R. Strong, J. Sun, D. Thomasova, L. Q. Ton, P. Topalis, Z. Tu, M. F. Unger, B. Walenz, A. Wang, J. Wang, M. Wang, X. Wang, K. J. Woodford, J. R. Wortman, M. Wu, A. Yao, E. M. Zdobnov, H. Zhang, Q. Zhao, S. Zhao, S. C. Zhu, I. Zhimulev, M. Coluzzi, A. della Torre, C. W. Roth, C. Louis, F. Kalush, R. J. Mural, E. W. Myers, M. D. Adams, H. O. Smith, S. Broder, M. J. Gardner, C. M. Fraser, E. Birney, P. Bork, P. T. Brey, J. C. Venter, J. Weissenbach, F. C. Kafatos, F. H. Collins and Hoffman SL. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, **298**: 129–149.

Hooper LV and Gordon JI. (2001). Commensal host-bacterial relationships in the gut. *Science*, **292**: 1115–1118.

Hosokawa T, Nikoh N, Koga R, Satô M, Tanahashi M, Meng XY and Fukatsu T. (2012). Reductive genome evolution, host-symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies. *The ISME Journal*, **6:** 577–587.

Hughes GL, Ren X, Ramirez JL, Sakamoto JM, Bailey JA, Jedlicka AE and Rasgon JL. (2011). *Wolbachia* infections in *Anopheles gambiae* cells: transcriptomic characterization of a novel host-symbiont interaction. *PLoS Pathogens*, **7**: e1001296.

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A and Yong SY. (2012). InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research*, **40**: D306–D312.

International *Glossina* Genome Initiative (IGGI). (2014). Genome Sequence of the Tsetse Fly (*Glossina morsitans*): Vector of African Trypanosomiasis. *Science*, **344**: 380–386.

Jurgenson CT, Begley TP and Ealick S E. (2009). The structural and biochemical foundations of thiamin biosynthesis. *Annual Review of Biochemistry*, **78**: 569–603.

Kanehisa M, Goto S, Sato Y, Furumichi M and Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, **40**: D109–14.

Kappmeier K and Nevill EM. (1999). Evaluation of coloured targets for the attraction of *Glossina brevipalpis* and *Glossina austeni* (Diptera: *Glossinidae*) in South Africa. *The Onderstepoort Journal of Veterinary Research*, **66**: 291–305.

Kennedy PGE. (2006). Diagnostic and neuropathogenesis issues in human African trypanosomiasis. *International Journal for Parasitology*, **36**: 505–512.

Kgori PM, Modo S and Torr SJ. (2006). The use of aerial spraying to eliminate tsetse from the Okavango Delta of Botswana. *Acta Tropica*, **99**: 184–199.

Lawhorn BG, Mehl RA and Begley T P. (2004). Biosynthesis of the thiamin pyrimidine: the reconstruction of a remarkable rearrangement reaction. *Organic and Biomolecular Chemistry*, **2**: 2538–2546.

Leak SGA. (1999). *Tsetse Biology and Ecology* (pp. 1–568). CABI Publishing.

Lee D, Redfern O and Orengo C. (2007). Predicting protein function from sequence and structure. *Nature Reviews. Molecular Cell Biology*, **8**: 995–1005.

Lehane M J. (2005). *The Biology of Blood-Sucking in Insects*. Cambridge University Press.

Lin H, Kwan AL and Dutcher SK. (2010). Synthesizing and salvaging NAD: lessons learned from *Chlamydomonas reinhardtii. PLoS Genetics*, **6**: e1001105

Lin S and Cronan JE. (2011). Closing in on complete pathways of biotin biosynthesis. *Molecular bioSystems*, **7**: 1811–1821.

Marx H, Mattanovich D and Sauer M. (2008). Overexpression of the riboflavin biosynthetic pathway in *Pichia pastoris. Microbial Cell Factories*, **7**: 23–30.

Mateos M, Castrezana SJ, Nankivell BJ, Estes AM, Markow T and Moran N. (2006). Heritable endosymbionts of *Drosophila. Genetics*, **174**: 363–376.

Medlock J, Atkins KE, Thomas DN, Aksoy S and Galvani AP. (2013). Evaluating Paratransgenesis as a Potential Control Strategy for African Trypanosomiasis. *PLoS Neglected Tropical Diseases*, **7**: e2374.

Minard G, Mavingui P and Moro CV. (2013). Diversity and function of bacterial microbiota in the mosquito holobiont. *Parasites and Vectors*, **6**: 146–151.

Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, Hradecky P and Lewis SE. (2002). Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biology*, **3**: RESEARCH0083.

Moran NA, McCutcheon JP and Nakabachi A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**: 165–190.

Mortazavi A, Williams BA, Mccue K, Schaeffer L and Wold B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, **5**: 1–8.

Nogge G. (1976). Sterility in tsetse flies (*Glossina morsitans* Westwood) caused by loss of symbionts. *Experientia*, **32**: 995–1006.

O'Neill SL, Gooding RH and Aksoy S. (1993). Phylogenetically distant symbiotic microorganisms reside in *Glossina* midgut and ovary tissues. *Medical and Veterinary Entomology*, **7**: 377–383.

Odiit M, Coleman PG, Liu WC, McDermott JJ, Fèvre EM, Welburn SC and Woolhouse MEJ. (2005). Quantifying the level of under-detection of Trypanosoma brucei rhodesiense sleeping sickness cases. *Tropical Medicine & International Health: TM & IH*, **10**: 840–849.

Pais R, Lohs C, Wu Y, Wang J and Aksoy S. (2008). The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Applied and Environmental Microbiology*, **74**: 5965–5974.

Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D and Yeates TO. (1999). Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **96**: 4285–4288.

Petersen FT, Meier R, Kutty SN and Wiegmann BM. (2007). The phylogeny and evolution of host choice in the *Hippoboscoidea* (Diptera) as reconstructed using four molecular markers. *Molecular Phylogenetics and Evolution*, **45**: 111–122.

Pinon V, Ravanel S, Douce R and Alban C. (2005). Biotin synthesis in plants. The first committed step of the pathway is catalyzed by a cytosolic 7-keto-8-aminopelargonic acid synthase. *Plant Physiology*, **139**: 1666–1676.

Politi C, Carrín G, Evans D, Kuzoe FA and Cattand PD. (1995). Cost-effectiveness analysis of alternative treatments of African gambiense trypanosomiasis in Uganda. *Health Economics*, **4**: 273–287.

Price DRG, Duncan RP, Shigenobu S and Wilson ACC. (2011). Genome expansion and differential expression of amino acid transporters at the aphid/*Buchnera* symbiotic interface. *Molecular Biology and Evolution*, **28**: 3113–3126.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C and Finn RD. (2012). The Pfam protein families database. *Nucleic Acids Research*, **40**: D290–D301.

Reizer A, Deutscher J, Saier MH and Reizer J. (1991). Analysis of the gluconate (gnt) operon of *Bacillus subtilis*. *Molecular Microbiology*, **5**: 1081–1089.

Rita R, Rebecca S, Jingwen W, Claudia L, Yi-neng W, Anna S and Serap A. (2012). Insight into the Transmission Biology and Species-Specific Functional Capabilities of Tsetse ( Diptera: *Glossinidae* ) Obligate Symbiont *Wigglesworthia. mBio*, **3**: 1–13.

Schramek N, Bracher A and Bacher A. (2001). Biosynthesis of riboflavin. Single turnover kinetic analysis of GTP cyclohydrolase II. *The Journal of Biological Chemistry*, **276**: 44157–44162.

Settembre E, Dorrestein PC, Zhai H, Chatterjee A, McLafferty FW, Begley TPand Ealick SE. (2004). Thiamin biosynthesis in *Bacillus subtilis*: structure of the thiazole synthase/sulfur carrier protein complex. *Biochemistry*, **43**: 11647–11657.

Shaw APM. (2009). Assessing the economics of animal trypanosomosis in Africa — history and current perspectives. *Journal of Veterinary Research*, **32**: 27–32.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y and Ishikawa H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, **407**: 81–86.

Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A and Xenarios I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, **41**: D344–D347.

Simarro PP, Diarra A, Ruiz Postigo J, Franco JR and Jannin JG. (2011). The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000-2009: the way forward. *PLoS Neglected Tropical Diseases*, **5**:e1007.

Sinkins S and O'Neill S. (2000). *Wolbachia* as a vehicle to modify insect populations. In: Handler A, James A. (Eds.), Insect Transgenesis. *CRC Press*, New York, pp. 271–287

Snyder AK, Deberry JW, Runyen-Janecky L and Rio RVM. (2010). Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: *Glossinidae*) symbionts. *Proceedings. Biological Sciences*, **277**: 2389–2397.

Suzuki Y and Brown G. (1974). The Biosynthesis of Folic Acid. *The Journal of Biological Chemistry*, **249**: 2405–2410.

Tambasco-Studart M, Titiz O, Raschle T, Forster G, Amrhein N, and Fitzpatrick T B. (2005). Vitamin B6 biosynthesis in higher plants. *Proceedings of the National Academy of Sciences of the United States of America*, **102**: 13687–13692.

Tanaka T, Tateno Y and Gojobori T. (2005). Evolution of vitamin B6 (pyridoxine) metabolism by gain and loss of genes. *Molecular Biology and Evolution*, **22**: 243–250.

Thompson JD, Higgins DG and Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-

specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**: 4673–4680.

Tian W and Skolnick J. (2003). How well is enzyme function conserved as a function of pairwise sequence identity? *Journal of Molecular Biology*, **333**: 863–882.

Tobe SS and Langley PA. (1978). Reproductive physiology of *Glossina*. *Annual Review of Entomology*, **23**: 283–307.

Toh H, Weiss BL Perkin SAH, Yamashita A, Oshima K, Hattori M and Aksoy S. (2006). Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Research*, **16**: 149–156.

Vale GA. (1993). Development of baits for tsetse flies (Diptera: *Glossinidae*) in Zimbabwe. *Journal of Medical Entomology*, **30**: 831–842.

Van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U and Moya A. (2003). Reductive genome evolution in *Buchnera aphidicola*. *Proceedings of the National Academy of Sciences of the United States of America*, **100**: 581–586.

Vreysen MJ B, Seck MT, Sall B and Bouyer J. (2013). Tsetse flies: their biology and control using area-wide integrated pest management approaches. *Journal of Invertebrate Pathology*, **112**:, 15–25.

Vreysen MJ, Saleh KM, Ali MY, Abdulla AM, Zhu ZR, Juma K G and Feldmann HU. (2000). *Glossina austeni* (Diptera: *Glossinidae*) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique. *Journal of Economic Entomology*, **93**: 123–135.

Wang J, Weiss BL and Aksoy S. (2013). Tsetse fly microbiota: form and function. *Frontiers in Cellular and Infection Microbiology*, **3**, 69–78.

Weiss BL, Mouchotte R, Rio RVM, Wu YN, Wu Z, Heddi A and Aksoy S. (2006). Interspecific transfer of bacterial endosymbionts between tsetse fly species: infection establishment and effect on host fitness. *Applied and Environmental Microbiology*, **72**: 7013–7021.

Weiss BL, Wang J and Aksoy S. (2011). Tsetse immune system maturation requires the presence of obligate symbionts in larvae. *PLoS Biology*, **9**: e1000619.

Welburn SC, Maudlin I and Ellis DS. (1987). *In vitro* cultivation of rickettsia-like-organisms from *Glossina* spp. *Annals of Tropical Medicine and Parasitology*, **81**: 331–335.

Wellde BT, Waema D, Chumo DA, Reardon MJ, Oloo F, Njogu AR and Mugutu S. (1989). Review of tsetse control measures taken in the Lambwe Valley in 1980-1984. *Annals of Tropical Medicine and Parasitology*, **83**: 119–125.

Wernegreen JJ. (2002). Genome evolution in bacterial endosymbionts of insects. *Nature Reviews. Genetics*, **3**: 850–861.

Werren, J. H. (1997). *Wolbachia* run amok. *Proceedings of the National Academy of Sciences of the United States of America*, **94**: 11154–11165.

Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD and Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. In *Methods in molecular biology* **112**: 531–52).

**Appendix 1: Grinding buffer**

Grinding buffer was prepared by mixing 10 ml of 0.1M of NaCl, 20 ml of 0.2M sucrose, 10 ml of 0.1M Tris (pH 9.2), 10 ml of 0.05M ethylenediaminetetraacetic acid (EDTA), 5 ml of 0.5% sodium dodecyl sulfate (SDS) and double distilled water to a final volume of 100ml.

**Appendix 2: Transformation media and reagents**

a) Isopropyl $\beta$-D-1-thiogalactopyranoside (IPTG) stock solution 0.1M contains 1.2g IPTG dissolved in 50ml distilled water and stored at 4°C

b) 5-bromo-4-chloro-3-indolyl-$\beta$-D-galactopyranoside (X-Gal) 2ml was prepared by dissolving 100g of 5-bromo 4-chloro 3-indolyl β-D-galactoside in 2ml of N, N'-dimethyl-formamide. It is stored at -20°C in an aluminium foil covered container.

c) Luria-Bertani (LB) medium: 10g Bacto-tryptone, 5g Bacto-yeast extract and 5g NaCl was dissolved in a litre of distilled water and its pH adjusted to pH 7 using NaOH. This was then autoclaved before use.

d) LB plates with ampicillin, IPTG and X-Gal: 15g agar was added to a litre of LB medium then autoclaved. This was allowed to cool to 50°C before ampicillin was added to a final concentration of 100μg/ml

e) Super Optimal broth with Catabolite repression (SOC) medium: 100ml was prepared by mixing 2g of bacto-tryptome, 0.5g bacto-yeast extract, 1ml 1M NaCl, 0.25ml 1M KCl, 1ml 2M $Mg^{2+}$ stock (filter sterilized), 1ml 2M glucose (filter sterilized) and distilled water to a final volume of 100ml.

f) 2M $Mg^{2+}$ stock: Prepared by mixing 20.3g $MgCl_2$ •$6H_2O$ and 24.65g $MgSO_4$ •$7H_2O$ and distilled water to a final volume of 100ml then filter sterilized.

## A) Nicotinamide

Asp →1→ iAsp →2→ QA →3→ NaMN →4→ NaAD →5→ NAD+ →7→ NADP

NA →9→ NAM →10→ NMN

8 (NaMN → NA), 4 (NMN → NAD+), 6
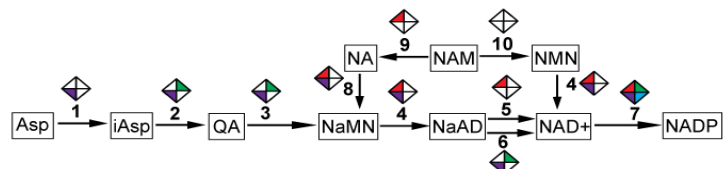
| Code | Enzyme | Code | Enzyme |
|---|---|---|---|
| 1 | L-aspartate oxidase | 6 | NH (3) dependent NAD synthetase |
| 2 | Quinolinate synthase A | 7 | NAD kinase |
| 3 | Quinolinate phosphoribosyl transferase | 8 | Nicotinate phosphoribosyl transferase |
| 4 | NMN adenylyltransferase | 9 | Nicotinamidase |
| 5 | Glutamine dependent NAD synthetase | 10 | Nicotinamide phosphoribosyltransferase |

KEY

| Asp | L-aspartate | NAM | Nicotinamide |
|---|---|---|---|
| iAsp | Iminoaspartate | NaMN | Nicotinatemononucleotide |
| QA | Quinolinate | NaAD | Nicotinate adenine dinucleotide |
| NADP | NAD phosphate | NAD | Nicotinamide adenine dinucleotide |
| NA | Nicotinic acid | NMN | Nicotinamide mononucleotide |

## B) Pantothenate

Val →1→ KiV →2→ KPa →3→ Pa →8→ Pan →9→ pPan →10→ pPanCy →11→ pPanN →12→ DCoA →13→ CoA

Ura →4→ dUra →5→ CAla →6→ Ala →7→ (Asp)
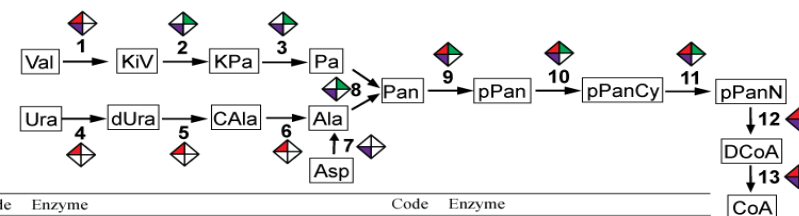
| Code | Enzyme | Code | Enzyme |
|---|---|---|---|
| 1 | Branched-chain-amino-acid aminotransferase | 8 | Pantothenate synthetase |
| 2 | Ketopantoate hydroxymethyltransferase | 9 | Pantothenate kinase |
| 3 | Ketopantoate reductase | 10 | Phosphopantothenate-cysteine ligase |
| 4 | Dihydropyrimidine dehydrogenase | 11 | Phosphopantothenoylcysteine decarboxylase |
| 5 | Dihydropyrimidinase | 12 | Phosphopantetheine adenylyltransferase |
| 6 | Beta-ureidopropionase | 13 | Dephospho-CoA kinase |
| 7 | Aspartate 1-decarboxylase | | |

KEY

| Val | L- valine | Asp | L- aspartate | CAla | N-carbamoyl β-alanine |
|---|---|---|---|---|---|
| Ura | Uracil | dUra | 5,6-dihydro-uracil | pPan | 4'-phosphopantothenate |
| Ala | β-alanine | CoA | Coenzyme A | pPanCy | 4'-phosphopantothenoyl cysteine |
| KPa | α- ketopantoate | Pan | Pantothenate | pPanN | 4'-phosphopantothenine |
| Pa | Pantoate | KiV | α- ketoisovalerate | DCoA | Dephospho coenzyme A |

## C) Thiamine

ThiF-SH

ThiI-SH →1→ ThiI-SSH → ThiS → ThiS-COSS-ThiF

Dx / dHG →2→ TPC

HMT → HMPP, HMP →6→ HMPP →7→ TMP →8→ TPP →10→ TTP

Az →5→, HMP, 9 (TMP → Thiamine), Thiamine →10→ TPP

11 (dHG/TPC), Tyrosine →3→, Glycine →4→

| Code | Enzyme | Code | Enzyme |
|---|---|---|---|
| 1 | Cysteine desulfurase | 6 | Phosphomethylpyrimidine kinase |
| 2 | Thiazole synthase | 7 | Thiamine-phosphate synthase |
| 3 | Thiamine biosynthesis protein ThiH | 8 | Thiamine-monophosphate kinase |
| 4 | Glycine oxidase | 9 | Non-specific phosphatase |
| 5 | Phosphomethylpyrimidine synthase | 10 | Thiamine pyrophosphokinase |
| | | 11 | Hydroxyethylthiazole kinase |

KEY

| Dx | 1-deoxy D-xylulose 5-phosphate | HMPP | Hydroxymethyl pyrimidine pyrophosphate |
|---|---|---|---|
| dHG | Dehydroglycine | TMP | Thiamine monophosphate |
| TPC | Thiazole phosphate carboxylate | TPP | Thiamine pyrophosphate |
| Az | 5-aminoimidazole | TTP | Thiamine triphosphate |
| HMP | Hydroxymethyl pyrimidine phosphate | HMT | Hydroxyethyl methylthiazole |

## D) Pyridoxine

EP →1→ PEA →2→ HPKB →3→ PHT →4→ PNP →6→ PLP →6→ PMP →8→ Pxm

dXP, PNP →5→, Pxn →7→, 8 (PNP → PLP)

| Code | Enzyme | Code | Enzyme |
|---|---|---|---|
| 1 | D-erythrose-4-phosphate dehydrogenase | 5 | Pyridoxine 5'-phosphate synthase |
| 2 | Erythronate-4-phosphate dehydrogenase | 6 | Pyridoxine-5'-phosphate oxidase |
| 3 | Phosphoserine aminotransferase | 7 | Pyridoxal phosphate phosphatase |
| 4 | 4-hydroxythreonine-4-phosphate dehydrogenase | 8 | Pyridoxal kinase |

KEY

| EP | D-erythrose 4-phosphate | PLP | Pyridoxal phosphate |
|---|---|---|---|
| PEA | 4-phospho D- erythronate | PMP | Pyridoxamine monophosphate |
| Pxm | Pyridoxamine | PHT | 4-phosphohydroxy L-threonine |
| Pxn | Pyridoxine | dXP | 1-deoxyxylulose 5-phosphate |
| PNP | Pyridoxine phosphate | HPKB | 3-hydroxy 4-phosphohydroxy α-ketobutyrate |

## E) Folate



| Code | Enzyme | Code | Enzyme |
|---|---|---|---|
| 1 | GTP cyclohydrolase I | 7 | P-aminobenzoic acid synthase |
| 2 | Alkaline phosphatase | 8 | Aminodeoxychorismate lyase |
| 3 | Dihydroneopterin pyrophosphatase | 9 | Dihydropteroate synthase |
| 4 | Non specific phosphatase | 10 | Folylpolyglutamate synthase |
| 5 | Dihydroneopterin aldolase | 11 | Dihydrofolate reductase |
| 6 | 2-amino-4-hydroxy-6-hydroxymethyl dihydropteridine pyrophosphokinase | | |

**KEY**

| | | | |
|---|---|---|---|
| GTP | Guanosine triphosphate | CM | Chorismate |
| dHPTP | Dihydroneopterin triphosphate | ADCM | Aminodeoxychorismate |
| dHT | Dihydroneopterin | PABA | Para aminobenzoic acid |
| dHPMP | Dihydroneopterin monophosphate | dHTe | Dihydropteroate |
| HMdHT | Hydroxymethyldihydropterin | dHf | Dihydrofolate |
| HMdHTP | Hydroxymethyldihydropterin pyrophosphate | tHF | Tetrahydrofolate |
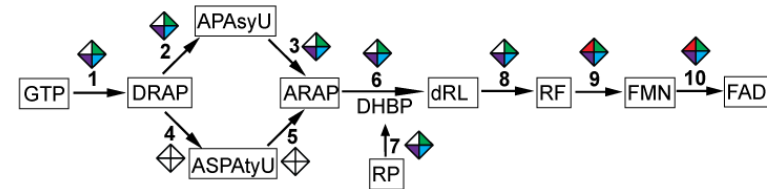
## F) Riboflavin



| Code | Enzyme | Code | Enzyme |
|---|---|---|---|
| 1 | GTP cyclohydrolase-2 | 6 | Lumazine synthase |
| 2 | DRAP deaminase | 7 | DHBP synthase |
| 3 | APAsyU reductase | 8 | Riboflavin synthase |
| 4 | DARP reductase | 9 | Riboflavin kinase |
| 5 | Bifunctional protein RIB2 | 10 | FAD synthase |

**KEY**

| | | | |
|---|---|---|---|
| GTP | Guanosine 5'-triphosphate | RP | Ribulose 5phosphate |
| ARAP | 5-amino 6-ribitylamino 2,4-(1H,3H)-pyrimidinedione | DHBP | 3,4-dihydroxy 2-butanone 4-phosphate |
| APAsyU | 5-amino 6-(5-phosphoD-ribosyl amino) uracil | dRL | 6,7-dimethyl 8-ribityl lumazine |
| APAtyU | 5-amino 6-(5-phospho D-ribityl amino) uracil | RF | Riboflavin |
| DRAP | 2,5-diamino 6-ribosylamino 4-(3H)-pyrimidinone 5'phosphate | FMN | Flavin mononucleotide |
| | | FAD | Flavin adenine dinucleotide |

## G) Biotin



| Code | Enzyme | Code | Enzyme |
|---|---|---|---|
| 1 | 8-amino-7-oxononanoate synthase | 3 | Dethiobiotin synthetase |
| 2 | Adenosylmethionine-8-amino-7-oxononanoate aminotransferase | 4 | Biotin synthase |
| | | 5 | Biotin protein ligase |

KEY

| | |
|---|---|
| PT | Pimelate thioester |
| Aon | 8-amino-7-oxononanoate |
| Don | 7,8-diaminononanoate |
| dTB | Dethiobiotin |
| Biotin-P | Biotin Protein |



*G. morsitans*   *W. glossinidia*
*S. glossinidius*   *W. pipientis*

**Figure A1.** Integrated pathways for vitamins and cofactors biosynthesis in *Glossina*, *Wigglesworthia*, *Sodalis* and *Wolbachia*. The arrows represent chemical reactions, numbers represent enzymes while products/ substrates are in squares. A coloured triangles indicate that the enzyme is present in the respective organism and absent if it is white (not coloured).

82

**Appendix 4: Table A1. Percentage identities between the sequenced and annotated genes**

| Gene IDs | Nucleotide sequence (%) | Amino acid sequence (%) |
|---|---|---|
| GMOY002354 | 97.0 (893/921) | 98.4 (301/306) |
| GMOY009051 | 97.1 (1024/1055) | 98.6 (287/291) |
| GMOY005148 | 96.1 (995/1035) | 97.2 (247/254) |
| GMOYdhfr2 | 97.0 (605/624) | 99.0 (185/187) |
| GMOY009056 | 96.0 (1537/1602) | 97.5 (161/165) |
| **Average identities** | **96.6** | **98.1** |

**Appendix 5: Alignment of isoenzymes**

**Alkaline phosphatases**

```
                1         10        20              30        40                              50
GMOY000067  MLFNY.KFYSQFFIIIIVFYCSHGLA..I.....DKSTTDDD..EERMHPLF.................NFKATER..SA
GMOY004796  MGSQI.LLYV...LIALLIDTI..RA..A....DHLKYGS.....................................S...KK
GMOY004885  MFYLI.YF...CVVSIILHCAKAATTSFP...PPIIRTLDT..TGQVGPEF..KTLDP.........AVPVDEA..TH
GMOY006875  MIAVRNDFYKFLFFVTLTLMFFRVLD..V....DCVAIGDP.YHKELF.D................AIGQKER..IL
GMOY007323  SCYVA......FVIIKYTLLLNRLVDIKFSHLLVNRSRSGTF.IHVKVRQDVNYSSIDGSEGTRMRNVDLNTLPSPSPSS
GMOY009926  MGTLR.KFIFPSLLIHLLLCTIESLE..Y.....VCSREDEDCFDRQMHPD..................LVRQGEE..L.

                            60        70        80        90
GMOY000067  ...........................RLEKRALPMLRFEYLKNRELDSNHWQELGQSILEKQLKSKST...
GMOY004796  ................................YTAKN........MSPEHTEYWNEKAQDILKQKLVQVQS...
GMOY004885  ................................KLAKANE.....DKKVEDSKFWFNVGLKQLEKIVNVAKNNDD
GMOY006875  ................................FYENRLT.......PEKERSTEFWLEEAQRYVKERVTG...S...
GMOY007323  TPMKFYHSNRFKIAATFLLLSLSILITGTCLAYYGKLANNSSATSGSFNLNNLPSEQVIWFEKGEAELLKALNR..K...
GMOY009926  ................................ERSIN.........P.LEYTNNYWLNLSREFIKQQMNR..Q...

             100       110       120       130       140       150       160
GMOY000067  .LNTNLAKNVIMFLGDGMSIPTLTAGRVYMG.......GEEKQFSFEEFPYMGLSKTYCVNTQVADSACTATAYLAGIKA
GMOY004796  .PITNNAKNIIMFLGDGMSVHTVTATRNFMG.......DSAKQISFEKFPYFGLSKTYCVDSQVADSACSATAYLGGVKG
GMOY004885  NSYKKKAKNIIIFVGDGMGVTTITAGRIYKGQYLKHGLGEEEKLSFDEFPYTALVKTYNVDKQVPDSAGTATAMFCGSKA
GMOY006875  .FNSQKAKNIILLGDGMSIATITAARILKGRQ.GCSGEESILTFDKFPFTGLVKTYCTNGQTPDSACTATAYLCGVKT
GMOY007323  .ENSRKARNVILFVADGMGPNTVTAARIYGF.......GEDGLMAWEEFPNMGLLKTYCANKLVPDSASTATALFSGVKT
GMOY009926  .PNTSKAKNIIFFLGDGMGLTTVAAARNLLG.......GTEQKLAFEEFPYSGLSKTYSVDRIVSDSACTATAYLCGVKA

             170       180       190       200       210       220       230       240
GMOY000067  NYATICLMAAVSLYDCAGQNNTAHHVSSLAAWAQAQGMATGLITNTRITHASPAGIFAHIANRNWEDDGHVLKDNG..D.
GMOY004796  NYGTICVNANVPRFDCMAGLNVGNWVESIAQWAQNAGMDTGLVTTARVTHASPAGVYAHTAQRGWENDAEILKSDC..S.
GMOY004885  HYGAICMDSTRSPNNST.....QGRLTSIMDWAQKENKRTGIVTTTRITHATPAATYARIYHRDWECDSKIPEQS....K
GMOY006875  NVMMLGCVNANVQNNNCSDSMNKEHQIDSIVQWAQKAGSTGLVTTTITHASPSAGYAHVANRLWECDNDINTMGAIIGP
GMOY007323  NYETSCVDASVDFSNCSASLNSLHQLPTIMSWAQDEGLHTGFVTTTRVTHATPSALYAHVPDRSWECESNMPESA....K
GMOY009926  IKGTICVNGQVNMGDCAAGIDRNNFVESIAKWAMDAGKSAGVVTTTRVTHASPAGVYANIADRDWENDSKVKEACG..TK

             250       260       270       280       290       300
GMOY000067  PKVCPDIAHQMIYSAVGRKLNVILGGGRKHFLPE.........L.......KSNGSRLDGRNLIQEWKSL...HGSTAHYVE
GMOY004796  PKVNVDIARQLVEWPVGRKLKVIMGGGRSNFRNV.TVIDE..D...GIAGQRYDGRDLIEYWIQDKRSQNVKANYVW
GMOY004885  .AHYIDIARQLVETSPGNKFNVVLGGGLQPLGAM.NPMEKRTTPWEGGTEKVCSRNDTLNLPEMWLKMGKEINDERAFVS
GMOY006875  GEECKDIAQQLITSEPGKNLNVILGGGMGKFLPK.HKVDA..F......GVKGERRDGLDIDHWQSN....NPQGIFVR
GMOY007323  MQGCTDIATQLVEGNVGQKINVIMGGGRQMLISNISGSEEDP..LD...HWAGYSKDGRHLIQKMMQKKQEQNRLYQVVQ
GMOY009926  SAKIEDIAYQLMHGEVGRKLSFIMGGGKQNFIDS.S..........YGNGKRSDGLDLIEYFQQE....HPDNVYVE

             310       320       330       340       350       360       370       380
GMOY000067  SREELFNLP.KSADKLLGLFAMDHLPYHLDSD..T...TTTPSLDEMVQVALHILKTQSQGKGYFLFVEGGRIDHAHHDSL
GMOY004796  NRKGLNSLNLNNTDYLLGLFSSSHCPFHGDLKRNNISTLVPSLSEMTESALKVLSK..KNNGFFLFVEGARIDMAHHETK
GMOY004885  TQEELKQLDLKKINSLMGLFRNNHMTYAIAKE..E...GEPSLKEMTEAAIDVLERGNKSKGYVLLVEGGRIDQGHHQNY
GMOY006875  NLEELEKVDFNQTKQLLGLFHSSHLEYYSKAKVE..RPKQPRLKDMTKAAIKTLEK..NSEGYFLFIEGGRIDHGHHETK
GMOY007323  NRDELMAINGTSVNYVLGIFANGHLKYDNELDMG..SRSMPSLKEMTVKALQILSS..RERGFLLVVEGGLIDQAHHRGK
GMOY009926  TRNELMQVNSLEANRVLGLYQDSHMLCHLETD.E..NTNQPTLEEMTIKAIEHLQH..NENGYFVFIECGLIDWLHHFNK

             390       400       410       420       430       440       450
GMOY000067  AMKALDETVEFDKAVQLAQLQTS...NDDTLIVVTSDHSHTMSVAGYSSRKNNIVGINNSQMGLDGLPYATISYANGPGY
GMOY004796  ARKSLEDTEEFAKAVDLTRKMFS...EEDTLIIVTSDHSHTMTISGYPYRDEDITGLAADQ.ADDDLPYTILSYANGPGF
GMOY004885  ARAALHEVYEFDMAIQAALEKTN...ENETLIIVTSDHSHTVLNGYPARGNDILGFANKP.GQDP..IYETLTYANGPGF
GMOY006875  AAYALDETLEFDEAIQTALDLTD...HRETLIVVTSDHSHTMTISGYPSRGNPILGLNERDLDSDGVPYSVLNYAIGPQ.
GMOY007323  AKKAINEVLAFNNAINASIQALNKEELEDTLIIVTADHSHTLTINGYPDRGTSVFGIAGKS.ITEGTPYTVLTYGTTRK.
GMOY009926  PRLALDETIEFSKAIQAARQITS...EEDTLIVVTADHSHAFSYGGYPQRKSDVFAPAPVS.GKDGKPYMPLNYANGPGF

             460       470       480       490       500       510       520       530
GMOY000067  TNNLLTKST..GFQRKNLHQVN..MDDKDYQFPSAVPLDSETHGGDDVGVFAKGPFAHLFTGVYEQNFIPHAIAYASCLG
GMOY004796  ENTYDKKG......RKRINPEE..TIDPNFQYLATVPLASETHGGDDVGVFASGPFAHYFSCTYEQSNIPAMMAKAANIG
GMOY004885  KDHLNLSATTENIWITNFTEKQ..RQSSTYRHLATLPLPDETHGGEDVALFAHGPGSSLVRGVFEQNYIAFVMSYAGCMG
GMOY006875  .QYMDS.N.....GKRLDLTN..FISDNMIFPSYIKALMGTHAGEDVAIFARGPQADLFSGTLQQNVIPHLMAYAACIG
GMOY007323  .GFQ.TD.D...KCLRKNPATDD..TESWEYTQQGAVNTEENNHGGSDVTIHATGAMSHLFHGVHEQSYVAHAISYALRIG
GMOY009926  SKFFDTES......GQRVDPSLNIVGNTDDEFPSTWPMQWETHGGDDVPVYASGPWSHLFTGVYEQNALPHFMAYAACLG

             540
GMOY000067  SQKTACSD.......................GLHPRR
GMOY004796  TFKSAAGVHARPT.......VMALFIIHF..FIILLKIR
GMOY004885  PAKDFDNSCQHRSGAGSLSISLLLFAAVSFVPYIHIFDN
GMOY006875  NGLHLCK......................QEDV
GMOY007323  RFRENSL.......................VKQLNE
GMOY009926  NGLTMC........................KT
```

# Dihydrofolate reductases

```
              1        10        20        30        40        50        60        70        80
              .        .         .         .         .         .         .         .         .
GMOY008444    INYKNFISKSNVKTIKMSKTPGVGTNSDATYSIQHGKVDTIDNDEGNRTSPLYVAFNETERPIEDATHNQVAMNSQNTKN
GMOYdhfr2     ...............................................................................

              90       100       110       120       130       140       150       160
              .        .         .         .         .         .         .         .
GMOY008444    LIDRKSDDPAVQANMKHGRLDIHDVQSKPEIEQICGDEKKTFLSKTRETTGVDFDKSDINAVVTDSSRQANENGGTIGGL
GMOYdhfr2     .............................................................................ML

              170       180       190       200       210       220       230
              .         .         .         .         .         .         .
GMOY008444    KFNLIVAVSKNLGIGLKGGLPWKLKSELKYFSQTKRVLDPTKRNVVVMGRKTYFGIPPSNRPLRDRLNIVLSTTLTK.S
GMOYdhfr2     KFNLIVAVSKNFGIGLKGGLPWELKSELRYFSEMTKRVFDSTKRNVVIMGRKTYFGIPLNNRPLRNRLNIVLSTTLNKVG

              240       250       260       270       280       290       300       310
              .         .         .         .         .         .         .         .
GMOY008444    DLPDEVLLQPNLEAAMKFFENNVLKNSIETVWIIGGAGVFKDAMASERCHRLYITQIQSNFECDVFLPAIPDDFQEVIT
GMOYdhfr2     ELPEEVLLQPNLEAAMKFLEDNNTLKSNIENIWIIGGASVFKEAMASKRCHRLYITEIQSSFESDVFLPTIPNDFEQIIP

              320       330       340
              .         .         .
GMOY008444    EPEIPQGMQAENGTNFVYKVFQK.R
GMOYdhfr2     GPEVPQTVQVENCICFRYKVLEKRE
```

# FAD synthases

```
              1        10        20        30        40        50        60        70        80
              .        .         .         .         .         .         .         .         .
GMOY000849    MEGETKDSEAKDSEAKDSKKRRKPKPKPKPKPPPEPIVYLGKTAKINAKLLKAKQRVPPQKKPEKKEEKKEGEEEPPPEK
GMOY008278    ...............................................................................

              90       100       110       120       130       140       150       160
              .        .         .         .         .         .         .         .
GMOY000849    KEKEEKKVQKSQKKVAKAKAEAKKKSEDAELARLVKRQELKAKYKEIIKVREKRKKKEEPEKKLCTITKGKAKAVVEEDK
GMOY008278    ...............................................................................

              170       180       190       200       210       220       230       240
              .         .         .         .         .         .         .         .
GMOY000849    KPGSGSGENLKDTIYHDMSMDIDPQEHRPKTEPELLLVPKYYKFEEIKSKVEAKEEPLTELFKKTFQMYKAQEVMLAFNG
GMOY008278    ................MSINSVLEENCLRANPEFILEAKY.SLEEIKSNIQIKEKSFEELCKKTFQMYKAEEVVLSFNG

              250       260       270       280       290       300       310       320
              .         .         .         .         .         .         .         .
GMOY000849    GKDCTVILHMLDLFFQKNHCLKHLKIPTLFITDPDGFPEVEEFVNDCAKLYNIELIKRPGTIKEALDEICKEKPLIRAIF
GMOY008278    GKDSTVVLHMLARFFQKDHNLKHLKILALFITDPDGFSEIDEFVDDCSKLYNIELIKMEGTIKQALERMCRERPLIRAVF

              330       340       350       360       370       380       390       400
              .         .         .         .         .         .         .         .
GMOY000849    MGSRRTDPHCQDLKVMQPTDPGWPPLMRINPILDWTCRDVWQYMYVYNVPYCILYQRGFTSIGNKKNSKPNPYLRVIEST
GMOY008278    MGSRRTDPHCQDLKTMQPTDPGWPALMRINPILEWTCRDIWQYIYVYDVAYCILYQKGFTSIGNKKNTKPNPYLQVAESQ

              410       420
              .         .
GMOY000849    TGRVLDYRPGHELLDNDNLERAGR.
GMOY008278    TGRVLNYRHAHELLDNDNLERAGRV
```

# Pantothenate kinase

```
              1        10        20        30        40        50        60        70
              .        .         .         .         .         .         .
GMOY000935    MKVRTCNHSTFNLKNYKFQAVSNNVNEIE.........KKPNGLAAVHNNNSKRKF.SRPSLKIVNPEKNNLKIKIAKE
GMOY006067    .............KSIKFLKICKNISVKHRIFLTIYLKHFLNMCELHCNVLVKDSNTYQPDTLDLNKDKEAA....AYW

              80       90        100       110       120       130       140       150
              .        .         .         .         .         .         .         .
GMOY000935    SKYKTNMVETNVFLNTRQQLKHETRAEASAVIEDELLMSMPWFGMDIGGTLTKLVYFEPKDITPDEQDQEAESLRSIRRY
GMOY006067    FPCFRDLIEKFAQRASESQRATDTSAEERANQFR.............SSYLSKLQ........EYQSQSESHKVLGIREL

              160       170       180       190       200       210
              .         .         .         .         .         .
GMOY000935    LTKNSAYGK.TGHRDTHLQMDNVEIRQRRGSLHFIRFQTTDM.....GNFLSLAKQ.......KGMAELVTTVCATGGGA
GMOY006067    LELNDAQLRLHGFRDPWQEQKKTENEEA.VPLLSARLQEIDAIESDRERWLELVRGVLAGNMFDWGAQAVTNILQE.DAS

              220       230       240       250       260       270       280       290
              .         .         .         .         .         .         .         .
GMOY000935    FKFENDFREQVNMKLAKFDELDTLIKGILFSEVQNRTECYYYENARDITKSEKRVYDFSQPYPFILVNVGSGVSILAVYG
GMOY006067    FGLNAALERIQKRPWLMDDLDAWLKRLQEEEIHKCAAIFVDNSGVDVVLG.......ILPFARELLKRGTKVLLCANTE

              300       310       320       330       340       350       360       370
              .         .         .         .         .         .         .         .
GMOY000935    PDNYKRVSGTSLGGGTFLGLCCLLTGCNTFEEAIQLATKGDNRKVDKLVRDIYGGDYDRFGLTGDLVASSFGQMHIRDKR
GMOY006067    PS........LNDVTSSELTAILDCCNCSCPV................................................

              380       390       400       410       420       430       440       450
              .         .         .         .         .         .         .         .
GMOY000935    ASVSREDLANATLVTITNNIGLIARMCALNEKMDKVVFVGNFLRVNPIAMKLLAYAM.EFWSNGTLKGLFLEHEGYFGAL
GMOY006067    .......................LKRACS..EK.RLLVFGNGQRGPCLDMRTLPQDLCEALKTYQTDLLIIEGMGRALHT

              460       470       480       490
              .         .         .         .
GMOY000935    GCLLQFNGEIA..AALNETSDKTAVNNPTPSALFSQPAPQR
GMOY006067    NLYARFNCETLKLAVVKNKWLAQRLGGDTFAVICKYEPLTS
```

## Table S1: Experimentally validated orthologs in B-vitamins and cofactors biosynthesis

| Entry | Organism | EC number | Pathway | Literature |
|-------|----------|-----------|---------|------------|
| P38891 | *Saccharomyces cerevisiae* | 2.6.1.42 | L-isoleucine biosynthesis | Colon M., Hernandez F., Lopez K., Quezada H., Gonzalez J., Lopez G., Aranda C., Gonzalez A. 2011 |
| P47176 | *Saccharomyces cerevisiae* | 2.6.1.42 | L-isoleucine biosynthesis | Colon M., Hernandez F., Lopez K., Quezada H., Gonzalez J., Lopez G., Aranda C., Gonzalez A. 2011 |
| P37653 | *Escherichia coli* | 2.4.1.12 | Bacterial cellulose biosynthesis | Zogaj X., Nimtz M., Rohde M., Bokranz W., Roemling U. 2001 |
| P12995 | *Escherichia coli* | 2.6.1.62 | Biotin biosynthesis | Stoner G.L., Eisenberg M.A. 1975 |
| P50277 | *Saccharomyces cerevisiae* | 2.6.1.62 | Biotin biosynthesis | Wu H, Ito K, Shimoi H. 2005. |
| P12996 | *Escherichia coli* | 2.8.1.6 | Biotin biosynthesis | Sanyal I., Cohen G., Flint D.H. 1994 |
| P12999 | *Escherichia coli* | 2.1.1.197 | Biotin biosynthesis | Del Campillo-Campbell A., Kayajanian G., Campbell A., Adhya S. 1967 |
| P13000 | *Escherichia coli* | 6.3.3.3 | Biotin biosynthesis | Eisenberg M.A., Krell K. 1969 |
| B0F481 | *Arabidopsis thaliana* | 6.3.3.3; 2.6.1.62 | Biotin biosynthesis | Patton D.A., Volrath S., Ward E.R. 1996 |
| P53556 | *Bacillus subtilis* | 2.3.1.47 | Biotin biosynthesis | Bower S., Perkins J.B., Yocum R.R., Howitt C.L., Rahaim P., Pero J. 1996 |
| P12998 | *Escherichia coli* | 2.3.1.47 | Biotin biosynthesis | Webster S.P., Alexeev D., Campopiano D.J., Watt R.M., Alexeeva M., Sawyer L., Baxter R.L. 2000 |
| P13001 | *Escherichia coli* | 3.1.1.85 | Biotin biosynthesis | Tomczyk N.H., Nettleship J.E., Baxter R.L., Crichton H.J., Webster S.P., Campopiano D.J. 2001 |
| P06709 | *Escherichia coli* | 6.3.4.15 | Biotin biosynthesis | Weaver L.H., Kwon K., Beckett D., Matthews B.W. 2001 |
| Q9UBR1 | *Homo sapiens* | 3.5.1.6 | Beta-alanine biosynthesis | Sakamoto T., Sakata S.F., Matsuda K., Horikawa Y., Tamaki N. 2001 |
| Q03941 | *Saccharomyces cerevisiae* | 2.7.1.24 | Coenzyme A biosynthesis | Olzhausen J., Schuebbe S., Schueller H.-J. 2009 |
| O48946 | *Arabidopsis thaliana* | 2.4.1.12 | Plant cellulose biosynthesis | Burn J.E., Hocart C.H., Birch R.J., Cork A.C., Williamson R.E. 2002 |
| P0ABQ0 | *Escherichia coli* | 4.1.1.36; 6.3.2.5 | Coenzyme A biosynthesis | Strauss E., Kinsland C., Ge Y., McLafferty F.W., Begley T.P. 2001 |
| Q96CD2 | *Homo sapiens* | 4.1.1.36 | Coenzyme A biosynthesis | Daugherty M., Polanuyer B., Farrell M., Scholle M., Lykidis A., de Crecy-Lagard V., Osterman A. 2002 |
| Q9ZPV8 | *Arabidopsis thaliana* | 2.7.7.3 | Coenzyme A biosynthesis | Kupke T., Hernandez-Acosta P., Culianez-Macia F.A. 2003 |
| P0A6I6 | *Escherichia coli* | 2.7.7.3 | Coenzyme A biosynthesis | Izard T. 2002 |
| P53332 | *Saccharomyces cerevisiae* | 2.7.7.3 | Coenzyme A biosynthesis | Olzhausen J., Schuebbe S., Schueller H.-J. 2009 |
| P0A6I9 | *Escherichia coli* | 2.7.1.24 | Coenzyme A biosynthesis | Mishra P.K., Park P.K., Drueckhammer D.G. 2001 |
| Q13057 | *Homo sapiens* | 2.7.7.3; 2.7.1.24 | Coenzyme A biosynthesis | Daugherty M., Polanuyer B., Farrell M., Scholle M., Lykidis A., de Crecy-Lagard V., Osterman A. 2002 |
| P37564 | *Bacillus subtilis* | 2.7.1.33 | Coenzyme A biosynthesis | Hong B.S., Yun M.K., Zhang Y.-M., Chohnan S., Rock C.O., White S.W., Jackowski S., Park H.-W., Leonardi R. 2006 |
| Q18677 | *Caenorhabditis elegans* | 3.5.2.2 | Coenzyme A biosynthesis | Takemoto T., Sasaki Y., Hamajima N., Goshima Y., Nonaka M., Kimura H. 2000 |
| P0ACl3 | *Escherichia coli* | 2.5.1.15 | Tetrahydrofolate biosynthesis | Swedberg G., Castensson S., Skold O. 1979 |
| Q12882 | *Homo sapiens* | 1.3.1.2 | Beta-alanine biosynthesis. | Lu Z.-H., Zhang R., Diasio R.B. 1992 |
| Q28943 | *Sus scrofa* | 1.3.1.2 | Beta-alanine biosynthesis. | Lohkamp B., Voevodskaya N., Lindqvist Y., Dobritzsch D. 2010 |
| Q86XF0 | *Homo sapiens* | 1.5.1.3 | Tetrahydrofolate biosynthesis | Anderson D.D., Quintero C.M., Stover P.J. 2011 |
| P0ABQ4 | *Escherichia coli* | 1.5.1.3 | Tetrahydrofolate biosynthesis | Bystroff C., Kraut J. 1991 |
| P00374 | *Homo sapiens* | 1.5.1.3 | Tetrahydrofolate biosynthesis | Anderson D.D., Quintero C.M., Stover P.J. 2011 |
| P0A9B6 | *Escherichia coli* | 1.2.1.72 | Pyridoxine 5'-phosphate biosynthesis | Yang Y., Zhao G., Man T.-K., Winkler M.E. 1998 |
| Q8NFF5 | *Homo sapiens* | 2.7.7.2 | FAD biosynthesis | Torchetti E.M., Brizio C., Colella M., Galluccio M., Giancaspero T.A., Indiveri C., Roberti M., Barile M. 2010 |
| P53848 | *Saccharomyces cerevisiae* | 4.1.2.25; 2.7.6.3; 2.5.1.15 | Tetrahydrofolate biosynthesis | Lawrence M.C., Iliades P., Fernley R.T., Berglez J., Pilling P.A., Macreadie I.G. 2005 |
| Q9SF23 | *Arabidopsis thaliana* | 4.1.2.25 | Tetrahydrofolate biosynthesis | Bauer S., Schott A.K., Illarionova V., Bacher A., Huber R., Fischer M. 2004 |
| P28823 | *Bacillus subtilis* | 4.1.2.25 | Tetrahydrofolate biosynthesis | Slock J., Stahly D.P., Han C.-Y., Six E.W., Crawford I.P. 1990 |
| P0ACl6 | *Escherichia coli* | 4.1.2.25 | Tetrahydrofolate biosynthesis | Haussmann C., Rohdich F., Schmidt E., Bacher A., Richter G. 1998 |

| Entry | Organism | EC number | Pathway | Literature |
|---|---|---|---|---|
| Q05932 | *Homo sapiens* | 6.3.2.17 | Tetrahydrofolylpolyglutamate biosynthesis. | Tomsho J.W., Moran R.G., Coward J.K. 2008 |
| Q08645 | *Saccharomyces cerevisiae* | 6.3.2.17 | Tetrahydrofolylpolyglutamate biosynthesis. | DeSouza L., Shen Y., Bognar A.L. 2000 |
| F4K2A1 | *Arabidopsis thaliana* | 6.3.2.17 | Tetrahydrofolylpolyglutamate biosynthesis. | Srivastava A.C., Ramos-Parra P.A., Bedair M., Robledo-Hernandez A.L., Tang Y., Sumner L.W., Diaz de la Garza R.I., Blancaflor E.B. 2011 |
| F4J2K2 | *Arabidopsis thaliana* | 6.3.2.17 | Tetrahydrofolylpolyglutamate biosynthesis. | Srivastava A.C., Ramos-Parra P.A., Bedair M., Robledo-Hernandez A.L., Tang Y., Sumner L.W., Diaz de la Garza R.I., Blancaflor E.B. 2011 |
| Q8W035 | *Arabidopsis thaliana* | 6.3.2.17 | Tetrahydrofolylpolyglutamate biosynthesis. | Srivastava A.C., Ramos-Parra P.A., Bedair M., Robledo-Hernandez A.L., Tang Y., Sumner L.W., Diaz de la Garza R.I., Blancaflor E.B. 2011 |
| P48596 | *Drosophila melanogaster* | 3.5.4.16 | 7,8-dihydroneopterin triphosphate biosynthesis | McLean J.R., Krishnakumar S., O'Donnell J.M. 1993 |
| P0A6T5 | *Escherichia coli* | 3.5.4.16 | 7,8-dihydroneopterin triphosphate biosynthesis | Lee S., Ahn C., Park E., Hwang D.S., Yim J. 2002 |
| P30793 | *Homo sapiens* | 3.5.4.16 | 7,8-dihydroneopterin triphosphate biosynthesis | Schoedon G., Redweik U., Curtius H.-C. 1989 |
| P94398 | *Bacillus subtilis* | 3.5.4.16 | 7,8-dihydroneopterin triphosphate biosynthesis | El Yacoubi B., Bonnett S., Anderson J.N., Swairjo M.A., Iwata-Reuyl D., de Crecy-Lagard V. 2006 |
| O31616 | *Bacillus subtilis* | 1.4.3.19 | Thiamine diphosphate biosynthesis. | Nishiya Y., Imanaka T. 1998 |
| P26281 | *Escherichia coli* | 2.7.6.3 | Tetrahydrofolate biosynthesis | Xiao B., Shi G., Chen X., Yan H., Ji X. 1999 |
| O31461 | *Bacillus subtilis* | 2.6.1.42 | L-isoleucine biosynthesis | Berger B.J., English S., Chan G., Knodel M.H. 2003 |
| P39576 | *Bacillus subtilis* | 2.6.1.42 | L-isoleucine biosynthesis | Berger B.J., English S., Chan G., Knodel M.H. 2003 |
| P0A6B7 | *Escherichia coli* | 2.8.1.7 | Thiamine biosynthesis | Lauhon C.T., Kambampati R. 2000 |
| Q8L3X9 | *Arabidopsis thaliana* | 2.3.1.41 | Fatty acid biosynthesis. | Ewald R., Kolukisaoglu U., Bauwe U., Mikkat S., Bauwe H. 2007 |
| P11458 | *Escherichia coli* | 2.5.1.72 | NAD(+) biosynthesis | Ollagnier-de Choudens S., Loiseau L., Sanakis Y., Barras F., Fontecave M. 2005 |
| P38032 | *Bacillus subtilis* | 1.4.3.16 | NAD(+) biosynthesis | Sun D., Setlow P.L. 1993 |
| P10902 | *Escherichia coli* | 1.4.3.16 | NAD(+) biosynthesis | Seifert J., Kunz N., Flachmann R., Laeufer A., Jany K.-D., Gassen H.G. 1990 |
| Q15274 | *Homo sapiens* | 2.4.2.19 | NAD(+) biosynthesis | Liu H., Woznica K., Catton G., Crawford A., Botting N., Naismith J.H. 2007 |
| P43619 | *Saccharomyces cerevisiae* | 2.4.2.19 | NAD(+) biosynthesis | Panozzo C., Nawara M., Suski C., Kucharczyka R., Skoneczny M., Becam A.-M., Rytka J., Herbert C.J. 2002 |
| P08164 | *Bacillus subtilis* | 6.3.1.5 | NAD(+) biosynthesis | Symersky J., Devedjiev Y., Moore K., Brouillette C., DeLucas L. 2002 |
| P18843 | *Escherichia coli* | 6.3.1.5 | NAD(+) biosynthesis | Jauch R., Humm A., Huber R., Wahl M.C. 2005 |
| Q6IA69 | *Homo sapiens* | 6.3.5.1 | NAD(+) biosynthesis | Hara N., Yamada K., Terashima M., Osago H., Shimoyama M., Tsuchiya M. 2003 |
| Q56YN3 | *Arabidopsis thaliana* | 2.7.1.23; 2.7.1.86 | NAD(+) biosynthesis | Berrin J.-G., Pierrugues O., Brutesco C., Alonso B., Montillet J.-L., Roby D., Kazmaier M. 2005 |
| O95544 | *Homo sapiens* | 2.7.1.23 | NAD(+) biosynthesis | Lerner F., Niere M., Ludwig A., Ziegler M. 2001 |
| Q9Y697 | *Homo sapiens* | 2.8.1.7 | Thiamine biosynthesis | Marelja Z., Stoecklein W., Nimtz M., Leimkuehler S. 2008 |
| Q93WX6 | *Arabidopsis thaliana* | 2.8.1.7; 4.4.1.16 | Thiamine biosynthesis | Pilon-Smits E.A.H., Garifullina G.F., Abdel-Ghany S., Kato S., Mihara H., Hale K.L., Burkhead J.L., Esaki N., Kurihara T., Pilon M. 2002 |
| Q9HAN9 | *Homo sapiens* | 2.7.7.1; 2.7.7.18 | NAD(+) biosynthesis | Sorci L., Cimadamore F., Scotti S., Petrelli R., Cappellacci L., Franchetti P., Orsomando G., Magni G. 2007 |
| Q9BZQ4 | *Homo sapiens* | 2.7.7.1; 2.7.7.18 | NAD(+) biosynthesis | Sorci L., Cimadamore F., Scotti S., Petrelli R., Cappellacci L., Franchetti P., Orsomando G., Magni G. 2007 |
| Q96T66 | *Homo sapiens* | 2.7.7.18; 2.7.7.1 | NAD(+) biosynthesis | Sorci L., Cimadamore F., Scotti S., Petrelli R., Cappellacci L., Franchetti P., Orsomando G., Magni G. 2007 |
| P39683 | *Saccharomyces cerevisiae* | 6.3.4.21 | NAD(+) biosynthesis | Panozzo C., Nawara M., Suski C., Kucharczyka R., Skoneczny M., Becam A.-M., Rytka J., Herbert C.J. 2002 |
| P0AFC0 | *Escherichia coli* | 3.6.1.- | NAD(+) biosynthesis | Gabelli S.B., Bianchet M.A., Xu W., Dunn C.A., Niu Z.-D., Amzel L.M., Bessman M.J. 2007 |
| Q9CA40 | *Arabidopsis thaliana* | 3.6.1.-; 3.6.1.55; 3.6.1.22 | Folate biosynthesis | Dobrzanska M., Szurmak B., Wyslouch-Cieszynska A., Kraszewska E. 2002 |
| Q9NWU1 | *Homo sapiens* | 2.3.1.41 | Fatty acid biosynthesis. | Zhang L., Joshi A.K., Hofmann J., Schweizer E., Smith S. 2005 |

| Entry | Organism | EC number | Pathway | Literature |
|---|---|---|---|---|
| P00903 | *Escherichia coli* | 2.6.1.85 | Tetrahydrofolate biosynthesis | Roux B., Walsh C.T. 1993 |
| P28820 | *Bacillus subtilis* | 2.6.1.85 | Tetrahydrofolate biosynthesis | Schadt H.S., Schadt S., Oldach F., Sussmuth R.D.(2009) |
| P28821 | *Bacillus subtilis* | 4.1.3.38 | Tetrahydrofolate biosynthesis | Slock J., Stahly D.P., Han C.-Y., Six E.W., Crawford I.P. 1990 |
| P28305 | *Escherichia coli* | 4.1.3.38 | Tetrahydrofolate biosynthesis | Green J.M., Nichols B.P. 1991 |
| P37254 | *Saccharomyces cerevisiae* | 2.6.1.85 | Tetrahydrofolate biosynthesis | Edman J.C., Goldstein A.L., Erbe J.G. 1993 |
| P31057 | *Escherichia coli* | 2.1.2.11 | (R)-pantothenate biosynthesis | Teller J.H., Powers S.G., Snell E.E. 1976 |
| P31663 | *Escherichia coli* | 6.3.2.1 | (R)-pantothenate biosynthesis | Miyatake K., Nakano Y., Kitaoka S. 1978 |
| P40459 | *Saccharomyces cerevisiae* | 6.3.2.1 | (R)-pantothenate biosynthesis | White W.H., Gunyuzlu P.L., Toyn J.H. 2001 |
| P0A790 | *Escherichia coli* | 4.1.1.11 | (R)-pantothenate biosynthesis | Cronan J.E. Jr. 1980 |
| P0A9J4 | *Escherichia coli* | 1.1.1.169 | (R)-pantothenate biosynthesis | Zheng R., Blanchard J.S. 2000 |
| Q04430 | *Saccharomyces cerevisiae* | 2.7.1.33 | Coenzyme A biosynthesis | Olzhausen J., Schuebbe S., Schueller H.-J. 2009 |
| O80448 | *Arabidopsis thaliana* | 4.3.3.6 | Pyridoxal 5'-phosphate biosynthesis. | Tambasco-Studart M., Titiz O., Raschle T., Forster G., Amrhein N., Fitzpatrick T.B. 2005 |
| Q9ZNR6 | *Arabidopsis thaliana* | 4.3.3.6 | Pyridoxal 5'-phosphate biosynthesis. | Tambasco-Studart M., Titiz O., Raschle T., Forster G., Amrhein N., Fitzpatrick T.B. 2005 |
| Q8L940 | *Arabidopsis thaliana* | 4.3.3.6 | Pyridoxal 5'-phosphate biosynthesis. | Tambasco-Studart M., Titiz O., Raschle T., Forster G., Amrhein N., Fitzpatrick T.B. 2005 |
| P19624 | *Escherichia coli* | 1.1.1.262 | Pyridoxine 5'-phosphate biosynthesis | Banks J., Cane D.E. 2004 |
| P05459 | *Escherichia coli* | 1.1.1.290 | Pyridoxine 5'-phosphate biosynthesis | Yang Y., Zhao G., Man T.-K., Winkler M.E. 1998 |
| P0AFI7 | *Escherichia coli* | 1.4.3.5 | B6 vitamer interconversion | di Salvo M., Yang E., Zhao G., Winkler M.E., Schirch V. 1998 |
| P0A794 | *Escherichia coli* | 2.6.99.2 | Pyridoxine 5'-phosphate biosynthesis | Laber B., Maurer W., Scharf S., Stepusin K., Schmidt F.S. 1999 |
| Q8W1X2 | *Arabidopsis thaliana* | 2.7.1.35 | B6 vitamer interconversion | Shi H., Zhu J.-K. 2002 |
| P39610 | *Bacillus subtilis* | 2.7.1.35 | Thiamine biosynthesis | Park J.-H., Burns K., Kinsland C., Begley T.P. 2004 |
| Q8TCD6 | *Homo sapiens* | 3.1.3.74 | Pyridoxine 5'-phosphate biosynthesis | Roberts S.J., Stewart A.J., Schmid R., Blindauer C.A., Bond S.R., Sadler P.J., Farquharson C. 2005 |
| Q96GD0 | *Homo sapiens* | 3.1.3.3; 3.1.3.74 | Vitamin B6 | Jang Y.M., Kim D.W., Kang T.-C., Won M.H., Baek N.-I., Moon B.J., Choi S.Y., Kwon O.-S. 2003 |
| P53184 | *Saccharomyces cerevisiae* | 3.5.1.19 | Nicotinate biosynthesis | Ghislain M., Talla E., Francois J.M. 2002 |
| Q6XQN6 | *Homo sapiens* | 6.3.4.21 | NAD(+) biosynthesis | Hara N., Yamada K., Shibata T., Osago H., Hashimoto T., Tsuchiya M. 2007 |
| Q9NVS9 | *Homo sapiens* | 1.4.3.5 | B6 vitamer interconversion | Musayev F.N., Di Salvo M.L., Ko T.-P., Schirch V., Safo M.K. 2003 |
| P00634 | *Escherichia coli* | 3.1.3.1 | Tetrahydrofolate biosynthesis | Kim E.E., Wyckoff H.W. 1991 |
| P11491 | *Saccharomyces cerevisiae* | 3.1.3.1; 3.1.7.6 | Tetrahydrofolate biosynthesis | Klionsky D.J., Emr S.D. 1989 |
| Q9HAB8 | *Homo sapiens* | 6.3.2.5 | Coenzyme A biosynthesis | Daugherty M., Polanuyer B., Farrell M., Scholle M., Lykidis A., de Crecy-Lagard V., Osterman A. 2002 |
| P40506 | *Saccharomyces cerevisiae* | 6.3.2.5 | Coenzyme A biosynthesis | Olzhausen J., Schuebbe S., Schueller H.-J. 2009 |
| Q9ZPY1 | *Arabidopsis thaliana* | 1.4.3.5 | B6 vitamer interconversion | Sang Y., Goertzen L.R., Tzou Y.-M., Locy R.D., Singh N.K. 2011 |
| Q12362 | *Saccharomyces cerevisiae* | 5.4.99.28; 3.5.4.26 | Riboflavin biosynthesis | Oltmanns O., Bacher A. 1972 |
| Q99258 | *Saccharomyces cerevisiae* | 4.1.99.12 | Riboflavin biosynthesis | Holt L.J., Tuch B.B., Villen J., Johnson A.D., Gygi S.P., Morgan D.O. 2009 |
| P50861 | *Saccharomyces cerevisiae* | 2.5.1.78 | Riboflavin biosynthesis | Moertl S., Fischer M., Richter G., Tack J., Weinkauf S., Bacher A. 1996 |
| O66747 | *Aquifex aeolicus* | 1.1.1.302 | Riboflavin biosynthesis. | Romisch-Margl W., Eisenreich W., Haase I., Bacher A., Fischer M. 2008 |
| P33312 | *Saccharomyces cerevisiae* | 1.1.1.302 | Riboflavin biosynthesis. | Oltmanns O., Bacher A. 1972 |
| P47924 | *Arabidopsis thaliana* | 4.1.99.12; 3.5.4.25 | Riboflavin biosynthesis | Hiltunen H.M., Illarionov B., Hedtke B., Fischer M., Grimm B. 2012 |
| P0A7I7 | *Escherichia coli* | 3.5.4.25 | Riboflavin biosynthesis | Foor F., Brown G.M. 1975 |
| P0A7J0 | *Escherichia coli* | 4.1.99.12 | Riboflavin biosynthesis | Richter G., Volk R., Krieger C., Laham H.-W., Roethlisberger U., Bacher A. 1992 |

| Entry | Organism | EC number | Pathway | Literature |
|---|---|---|---|---|
| P17618 | *Bacillus subtilis* | 3.5.4.26; 1.1.1.193 | Riboflavin biosynthesis | Richter G., Fischer M., Krieger C., Eberhardt S., Luttgen H., Gerstenschlager I., Bacher A. 1997 |
| P25539 | *Escherichia coli* | 3.5.4.26; 1.1.1.193 | Riboflavin biosynthesis | Richter G., Fischer M., Krieger C., Eberhardt S., Luttgen H., Gerstenschlager I., Bacher A. 1997 |
| P0AG40 | *Escherichia coli* | 2.7.1.26; 2.7.7.2 | FAD biosynthesis | Kamio Y., Lin C.-K., Regue M., Wu H.C. 1985 |
| Q9STY4 | *Arabidopsis thaliana* | 1.1.1.193 | Riboflavin biosynthesis | Hasnain G., Frelin O., Roje S., Ellens K.W., Ali K., Guan J.C., Garrett T.J., de Crecy-Lagard V., Gregory J.F. III, McCarty D.R., Hanson A.D. 2013 |
| P94465 | *Bacillus subtilis* | 2.7.1.26 | Riboflavin biosynthesis | Solovieva I.M., Kreneva R.A., Errais Lopes L., Perumov D.A. 2005 |
| Q969G6 | *Homo sapiens* | 2.7.1.26 | FMN biosynthesis | Yazdanpanah B., Wiegmann K., Tchikov V., Krut O., Pongratz C., Schramm M., Kleinridders A., Wunderlich T., Kashkar H., Utermoehlen O., Bruening J.C., Schuetze S., Kroenke M. 2009 |
| Q03778 | *Saccharomyces cerevisiae* | 2.7.1.26 | FMN biosynthesis | Santos M.A., Jimenez A., Revuelta J.L. 2000 |
| P16440 | *Bacillus subtilis* | 2.5.1.9 | Riboflavin biosynthesis | Bacher A., Baur R., Eggers U., Harders H.D., Otto M.K., Schnepple H. 1980 |
| P0AFU8 | *Escherichia coli* | 2.5.1.9 | Riboflavin biosynthesis | Eberhardt S.M.R., Richter G., Gimbel W., Werner T., Bacher A. 1996 |
| P38145 | *Saccharomyces cerevisiae* | 2.5.1.9 | Riboflavin biosynthesis | Santos M.A., Garcia-Ramirez J.J., Revuelta J.L. 1995 |
| P11998 | *Bacillus subtilis* | 2.5.1.78 | Riboflavin biosynthesis | Fischer M., Haase I., Kis K., Meining W., Ladenstein R., Cushman M., Schramek N., Huber R., Bacher A. 2003 |
| P61714 | *Escherichia coli* | 2.5.1.78 | Riboflavin biosynthesis | Moertl S., Fischer M., Richter G., Tack J., Weinkauf S., Bacher A. 1996 |
| P23721 | *Escherichia coli* | 2.6.1.52 | L-serine biosynthesis | Lam H.-M., Winkler M.E. 1990 |
| Q03148 | *Saccharomyces cerevisiae* | 4.3.3.6 | Pyridoxal 5'-phosphate biosynthesis. | Rodriguez-Navarro S., Llorente B., Rodriguez-Manzaneque M.T., Ramne A., Uber G., Marchesan D., Dujon B., Herrero E., Sunnerhagen P., Perez-Ortin J.E. 2002 |
| P43545 | *Saccharomyces cerevisiae* | 4.3.3.6 | Pyridoxal 5'-phosphate biosynthesis. | Rodriguez-Navarro S., Llorente B., Rodriguez-Manzaneque M.T., Ramne A., Uber G., Marchesan D., Dujon B., Herrero E., Sunnerhagen P., Perez-Ortin J.E. 2002 |
| P77444 | *Escherichia coli* | 2.8.1.7; 4.4.1.16 | Iron-sulfur cluster biosynthesis. | Outten F.W., Wood M.J., Munoz F.M., Storz G. 2003 |
| Q08224 | *Saccharomyces cerevisiae* | 2.7.1.49; 2.7.4.7 | Thiamine diphosphate biosynthesis | Llorente B., Fairhead C., Dujon B. 1999 |
| Q08975 | *Saccharomyces cerevisiae* | 2.7.1.49; 2.7.4.7 | Thiamine diphosphate biosynthesis | Llorente B., Fairhead C., Dujon B. 1999 |
| O82392 | *Arabidopsis thaliana* | 4.1.99.17 | Thiamine diphosphate biosynthesis | Raschke M., Buerkle L., Mueller N., Nunes-Nesi A., Fernie A.R., Arigoni D., Amrhein N., Fitzpatrick T.B. 2007 |
| P45740 | *Bacillus subtilis* | 4.1.99.17 | Thiamine diphosphate biosynthesis | Zhang Y., Begley T.P. 1997 |
| P30136 | *Escherichia coli* | 4.1.99.17 | Thiamine diphosphate biosynthesis | Lawhorn B.G., Gerdes S.Y., Begley T.P. 1994 |
| O31620 | *Bacillus subtilis* | 2.7.1.49; 2.7.4.7 | Thiamine diphosphate biosynthesis | Park J.-H., Burns K., Kinsland C., Begley T.P. 2004 |
| P76422 | *Escherichia coli* | 2.7.1.49; 2.7.4.7 | Thiamine diphosphate biosynthesis | Mizote T., Tsuda M., Smith D.D.S., Nakayama H., Nakazawa T. 1999 |
| P39594 | *Bacillus subtilis* | 2.5.1.3 | Thiamine diphosphate biosynthesis | Zhang Y., Taylor S.V., Chiu H.-J., Begley T.P. 1997 |
| P30137 | *Escherichia coli* | 2.5.1.3 | Thiamine diphosphate biosynthesis | Lawhorn B.G., Gerdes S.Y., Begley T.P. 1994 |
| P30138 | *Escherichia coli* | 2.7.7.73 | Thiamine diphosphate biosynthesis. | Taylor S.V., Kelleher N.L., Kinsland C., Chiu H.-J., Costello C.A., Backstrom A.D., McLafferty F.W., Begley T.P. 1998 |
| O31618 | *Bacillus subtilis* | 2.8.1.10 | Thiamine diphosphate biosynthesis. | Park J.-H., Dorrestein P.C., Zhai H., Kinsland C., McLafferty F.W., Begley T.P. 2003 |
| P30139 | *Escherichia coli* | 2.8.1.10 | Thiamine diphosphate biosynthesis. | Leonardi R., Fairhurst S.A., Kriek M., Lowe D.J., Roach P.L. 2003 |
| P30140 | *Escherichia coli* | 4.1.99.19 | Thiamine diphosphate biosynthesis. | Leonardi R., Fairhurst S.A., Kriek M., Lowe D.J., Roach P.L. 2003 |

| Entry | Organism | EC number | Pathway | Literature |
|-------|----------|-----------|---------|------------|
| P0AGG0 | *Escherichia coli* | 2.7.4.16 | Thiamine diphosphate biosynthesis | Nishino H. 1972 |
| O31617 | *Bacillus subtilis* | ThiS | Thiamine diphosphate biosynthesis. | Dorrestein P.C., Zhai H., McLafferty F.W., Begley T.P. 2004 |
| O32583 | *Escherichia coli* | ThiS | Thiamine diphosphate biosynthesis. | Taylor S.V., Kelleher N.L., Kinsland C., Chiu H.-J., Costello C.A., Backstrom A.D., McLafferty F.W., Begley T.P. 1998 |
| Q9H3S4 | *Homo sapiens* | 2.7.6.2 | Thiamine diphosphate biosynthesis | Nosaka K., Onozuka M., Kakazu N., Hibi S., Nishimura H., Nishino H., Abe T. 2001 |
| Q5M731 | *Arabidopsis thaliana* | 2.5.1.3; 2.7.1.49 | Thiamine diphosphate biosynthesis | Komeda Y., Tanaka M., Nishimune T. 1988 |
| P21829 | *Escherichia coli* | 3.1.3.74 | Pyridoxine 5'-phosphate biosynthesis | Maupin-Furlow J.A., Rosentel J.K., Lee J.H., Deppenmeier U., Gunsalus R.P., Shanmugam K.T. 1995 |

## Table S2: Domain organization of orthologs

| Pathway | Organism | Seq ID | Domain name | E-value | Predicted active site residues |
|---|---|---|---|---|---|
| Thiamine | *Drosophilla* | Q9VKD3 | Aminotran_5 | 1.30E-93 | [386,386] |
| Thiamine | *Tribolium* | D6WLY0 | Aminotran_5 | 4.30E-90 | [369,369] |
| Thiamine | *Anopheles* | Q7PW45 | Aminotran_5 | 6.70E-92 | [370,370] |
| Thiamine | *Wigglesworthia* | H6Q4V9 | Aminotran_5 | 5.50E-140 | [365,365] |
| Thiamine | *Wigglesworthia* | H6Q534 | Aminotran_5 | 1.20E-79 | [328,328] |
| Thiamine | *Sodalis* | Q2NT15 | Aminotran_5 | 1.10E-150 | [365,365] |
| Thiamine | *Sodalis* | Q2NS31 | Aminotran_5 | 6.20E-81 | [328,328] |
| Thiamine | *Wolbachia* | Q73H67 | Aminotran_5 | 7.50E-64 | [323,323] |
| Thiamine | *Wolbachia* | Q73GF8 | Aminotran_5 | 1.30E-85 | [339,339] |
| Thiamine | *Ecoli* | P77444 | Aminotran_5 | 6.20E-159 | [364,364] |
| Thiamine | *Ecoli* | P0A6B7 | Aminotran_5 | 9.10E-93 | [328,328] |
| Thiamine | *Bacillus* | O34599 | Aminotran_5 | 1.30E-87 | [325,325] |
| Thiamine | *Bacillus* | P38033 | Aminotran_5 | 1.40E-92 | [325,325] |
| Thiamine | *Bacillus* | O34874 | Aminotran_5 | 3.80E-78 | [324,324] |
| Thiamine | *Bacillus* | O32164 | Aminotran_5 | 2.80E-163 | [361,361] |
| Thiamine | *Arabidopsis* | Q93WX6 | Aminotran_5 | 3.30E-144 | [418,418] |
| Thiamine | *Arabidopsis* | O49543 | Aminotran_5 | 1.30E-89 | [377,377] |
| Thiamine | *Sorghum* | C5Y0K0 | Aminotran_5 | 2.60E-130 | [330,330] |
| Thiamine | *Yeast* | P25374 | Aminotran_5 | 2.40E-104 | [421,421] |
| Thiamine | *Human* | Q9Y697 | Aminotran_5 | 4.50E-94 | [381,381] |
| Thiamine | *Yeast* | Q08224 | TENA_THI-4 | 6.20E-75 | [468] |
| Thiamine | *Yeast* | Q08975 | Phos_pyr_kin | 1.40E-80 | - |
| Thiamine | *Yeast* | Q08975 | TENA_THI-4 | 4.40E-68 | [468] |
| Thiamine | *Ecoli* | P30138 | ThiF | 1.20E-43 | - |
| Thiamine | *Ecoli* | P30138 | MoeZ_MoeB | 7.00E-29 | [184] |
| Thiamine | *Bacillus* | O31619 | ThiF | 1.00E-41 | - |
| Thiamine | *Bacillus* | O31619 | MoeZ_MoeB | 3.40E-21 | [181] |
| Thiamine | *Wigglesworthia* | H6Q4N2 | ThiS | 6.10E-08 | - |
| Thiamine | *Ecoli* | O32583 | ThiS | 1.00E-17 | - |
| Thiamine | *Bacillus* | O31617 | ThiS | 2.30E-10 | - |
| Thiamine | *Wigglesworthia* | H6Q4N4 | Radical_SAM | 4.00E-07 | - |
| Thiamine | *Wigglesworthia* | H6Q4N4 | BATS | 1.10E-20 | - |
| Thiamine | *Sodalis* | Q2NWR4 | Radical_SAM | 1.30E-09 | - |
| Thiamine | *Ecoli* | P30140 | Radical_SAM | 2.00E-08 | - |
| Thiamine | *Ecoli* | P30140 | BATS | 5.70E-21 | - |
| Thiamine | *Bacillus* | O31616 | DAO | 4.40E-73 | - |
| Thiamine | *Wigglesworthia* | H6Q4N3 | ThiG | 3.30E-104 | [95] |
| Thiamine | *Ecoli* | P30139 | ThiG | 1.70E-109 | [95] |
| Thiamine | *Bacillus* | O31618 | ThiG | 1.10E-109 | [98] |
| Thiamine | *Wigglesworthia* | H6Q4N1 | TMP-TENI | 1.60E-50 | - |
| Thiamine | *Ecoli* | P30137 | TMP-TENI | 7.60E-57 | - |
| Thiamine | *Bacillus* | P39594 | TMP-TENI | 2.00E-61 | - |
| Thiamine | *Yeast* | P41835 | TMP-TENI | 2.20E-58 | - |
| Thiamine | *Yeast* | P41835 | HK | 2.40E-101 | - |
| Thiamine | *Wigglesworthia* | H6Q5T5 | AIRS | 1.90E-19 | - |
| Thiamine | *Wigglesworthia* | H6Q5T5 | AIRS_C | 2.10E-07 | - |
| Thiamine | *Sodalis* | Q2NV96 | AIRS | 6.10E-18 | - |
| Thiamine | *Sodalis* | Q2NV96 | AIRS_C | 4.10E-07 | - |
| Thiamine | *Ecoli* | P0AGG0 | AIRS | 3.10E-23 | - |
| Thiamine | *Ecoli* | P0AGG0 | AIRS_C | 1.80E-09 | - |
| Thiamine | *Bacillus* | O05514 | AIRS | 1.70E-22 | - |
| Thiamine | *Bacillus* | O05514 | AIRS_C | 8.50E-07 | - |
| Thiamine | *Drosophilla* | Q8T4A5 | TPK_catalytic | 2.10E-42 | - |
| Thiamine | *Drosophilla* | Q8T4A5 | TPK_B1_binding | 5.20E-19 | - |
| Thiamine | *Anopheles* | Q7QD56 | TPK_catalytic | 2.80E-28 | - |
| Thiamine | *Anopheles* | Q7QD56 | TPK_B1_binding | 2.40E-17 | - |
| Thiamine | *Bacillus* | O34664 | TPK_catalytic | 1.70E-31 | - |
| Thiamine | *Bacillus* | O34664 | TPK_B1_binding | 5.80E-18 | - |
| Thiamine | *Yeast* | P35202 | TPK_catalytic | 1.90E-17 | - |
| Thiamine | *Yeast* | P35202 | TPK_B1_binding | 8.80E-14 | - |
| Thiamine | *Human* | Q9H3S4 | TPK_catalytic | 9.90E-36 | - |
| Thiamine | *Human* | Q9H3S4 | TPK_B1_binding | 1.10E-25 | - |
| Riboflavin | *Wigglesworthia* | H6Q580 | DHBP_synthase | 3.60E-78 | - |
| Riboflavin | *Sodalis* | Q2NWD7 | DHBP_synthase | 2.90E-79 | - |
| Riboflavin | *Wolbachia* | Q73HB4 | DHBP_synthase | 4.00E-74 | - |
| Riboflavin | *Ecoli* | P0A7J0 | DHBP_synthase | 7.30E-80 | - |
| Riboflavin | *Bacillus* | P17620 | DHBP_synthase | 2.80E-84 | - |
| Riboflavin | *Bacillus* | P17620 | GTP_cyclohydro2 | 4.90E-80 | [330] |
| Riboflavin | *Arabidopsis* | Q6NLQ7 | DHBP_synthase | 2.50E-83 | - |
| Riboflavin | *Arabidopsis* | Q6NLQ7 | GTP_cyclohydro2 | 5.00E-51 | [429] |
| Riboflavin | *Arabidopsis* | F4KJA1 | DHBP_synthase | 1.50E-62 | - |

| Pathway | Organism | Seq ID | Domain name | E-value | Predicted active site residues |
|---|---|---|---|---|---|
| Riboflavin | *Arabidopsis* | F4KJA1 | GTP_cyclohydro2 | 2.40E-74 | [467] |
| Riboflavin | *Arabidopsis* | P47924 | DHBP_synthase | 1.00E-82 | - |
| Riboflavin | *Arabidopsis* | P47924 | GTP_cyclohydro2 | 1.40E-72 | [459] |
| Riboflavin | *Sorghum* | C5XVM3 | DHBP_synthase | 3.00E-83 | - |
| Riboflavin | *Sorghum* | C5XVM3 | GTP_cyclohydro2 | 7.60E-74 | [468] |
| Riboflavin | *Sorghum* | C5YJR6 | DHBP_synthase | 5.90E-84 | - |
| Riboflavin | *Sorghum* | C5YJR6 | GTP_cyclohydro2 | 8.70E-73 | [444] |
| Riboflavin | *Sorghum* | C5YZA3 | DHBP_synthase | 4.50E-67 | - |
| Riboflavin | *Sorghum* | C5YZA3 | GTP_cyclohydro2 | 4.60E-75 | [510] |
| Riboflavin | *Yeast* | Q99258 | DHBP_synthase | 1.80E-78 | - |
| Riboflavin | *Wigglesworthia* | H6Q4V2 | GTP_cyclohydro2 | 3.90E-64 | [128] |
| Riboflavin | *Sodalis* | Q2NT40 | GTP_cyclohydro2 | 1.20E-66 | [128] |
| Riboflavin | *Wolbachia* | Q73IY9 | GTP_cyclohydro2 | 5.10E-58 | [291] |
| Riboflavin | *Ecoli* | P0A7I7 | GTP_cyclohydro2 | 1.50E-69 | [128] |
| Riboflavin | *Yeast* | P38066 | GTP_cyclohydro2 | 5.50E-52 | [233] |
| Riboflavin | *Wigglesworthia* | H6Q5T9 | dCMP_cyt_deam_1 | 8.40E-19 | [58] |
| Riboflavin | *Wigglesworthia* | H6Q5T9 | RibD_C | 1.10E-39 | - |
| Riboflavin | *Sodalis* | Q2NV99 | dCMP_cyt_deam_1 | 2.20E-21 | [53] |
| Riboflavin | *Sodalis* | Q2NV99 | RibD_C | 2.10E-52 | - |
| Riboflavin | *Wolbachia* | Q73H62 | dCMP_cyt_deam_1 | 4.40E-20 | [52] |
| Riboflavin | *Wolbachia* | Q73H62 | RibD_C | 2.50E-40 | - |
| Riboflavin | *Ecoli* | P25539 | dCMP_cyt_deam_1 | 2.10E-24 | [52] |
| Riboflavin | *Ecoli* | P25539 | RibD_C | 5.70E-64 | - |
| Riboflavin | *Bacillus* | P17618 | dCMP_cyt_deam_1 | 2.40E-22 | [51] |
| Riboflavin | *Bacillus* | P17618 | RibD_C | 3.00E-65 | - |
| Riboflavin | *Arabidopsis* | Q8GWP5 | dCMP_cyt_deam_1 | 6.70E-21 | [123] |
| Riboflavin | *Sorghum* | C5X102 | dCMP_cyt_deam_1 | 1.60E-20 | [101] |
| Riboflavin | *Wigglesworthia* | H6Q5T8 | DMRL_synthase | 9.10E-44 | - |
| Riboflavin | *Sodalis* | Q2NV98 | DMRL_synthase | 5.30E-56 | - |
| Riboflavin | *Wolbachia* | P61729 | DMRL_synthase | 8.80E-42 | - |
| Riboflavin | *Ecoli* | P61714 | DMRL_synthase | 2.70E-57 | - |
| Riboflavin | *Bacillus* | P11998 | DMRL_synthase | 9.70E-59 | - |
| Riboflavin | *Arabidopsis* | O80575 | DMRL_synthase | 2.90E-51 | - |
| Riboflavin | *Sorghum* | C5Y9F9 | DMRL_synthase | 2.30E-51 | - |
| Riboflavin | *Yeast* | P50861 | DMRL_synthase | 9.30E-44 | - |
| Riboflavin | *Wigglesworthia* | H6Q4Y2 | Lum_binding | 1.70E-24 | - |
| Riboflavin | *Wigglesworthia* | H6Q4Y2 | Lum_binding | 8.80E-21 | - |
| Riboflavin | *Sodalis* | Q2NT11 | Lum_binding | 6.20E-25 | - |
| Riboflavin | *Sodalis* | Q2NT11 | Lum_binding | 4.70E-27 | - |
| Riboflavin | *Wolbachia* | Q73IM7 | Lum_binding | 1.50E-17 | - |
| Riboflavin | *Wolbachia* | Q73IM7 | Lum_binding | 9.60E-17 | - |
| Riboflavin | *Ecoli* | P0AFU8 | Lum_binding | 5.80E-30 | - |
| Riboflavin | *Ecoli* | P0AFU8 | Lum_binding | 2.20E-29 | - |
| Riboflavin | *Bacillus* | P16440 | Lum_binding | 2.00E-30 | - |
| Riboflavin | *Bacillus* | P16440 | Lum_binding | 3.00E-26 | - |
| Riboflavin | *Arabidopsis* | Q84MD8 | HAD_2 | 2.70E-25 | [17] |
| Riboflavin | *Arabidopsis* | Q84MD8 | Flavokinase | 3.00E-36 | - |
| Riboflavin | *Sorghum* | C5X020 | HAD_2 | 9.30E-27 | [17] |
| Riboflavin | *Sorghum* | C5X020 | Flavokinase | 3.10E-34 | - |
| Riboflavin | *Yeast* | Q03778 | Flavokinase | 1.40E-30 | - |
| Riboflavin | *Human* | Q969G6 | Flavokinase | 1.50E-39 | - |
| Riboflavin | *Drosophilla* | Q9VJY1 | PAPS_reduct | 3.30E-21 | - |
| Riboflavin | *Drosophilla* | Q8IR76 | PAPS_reduct | 4.20E-10 | - |
| Riboflavin | *Drosophilla* | Q8IR76 | PAPS_reduct | 1.00E-20 | - |
| Riboflavin | *Tribolium* | D6WDE0 | MoCF_biosynth | 4.30E-07 | - |
| Riboflavin | *Tribolium* | D6WDE0 | PAPS_reduct | 2.80E-26 | - |
| Riboflavin | *Anopheles* | F5HKN2 | PAPS_reduct | 1.10E-32 | - |
| Riboflavin | *Yeast* | P38913 | PAPS_reduct | 1.70E-41 | - |
| Riboflavin | *Human* | Q8NFF5 | MoCF_biosynth | 1.70E-28 | - |
| Riboflavin | *Human* | Q8NFF5 | PAPS_reduct | 6.20E-06 | - |
| Riboflavin | *Human* | Q8NFF5 | PAPS_reduct | 3.10E-17 | - |
| Riboflavin | *Human* | P30043 | NAD_binding_10 | 1.90E-32 | - |
| VitaminB6 | *Ecoli* | P0A9B6 | Gp_dh_N | 2.50E-52 | [155] |
| VitaminB6 | *Ecoli* | P0A9B6 | Gp_dh_C | 5.80E-53 | - |
| VitaminB6 | *Wigglesworthia* | H6Q509 | 2-Hacid_dh | 6.20E-16 | - |
| VitaminB6 | *Wigglesworthia* | H6Q509 | 2-Hacid_dh_C | 1.10E-32 | [254,237,208] |
| VitaminB6 | *Wigglesworthia* | H6Q509 | DUF3410 | 1.90E-16 | - |
| VitaminB6 | *Sodalis* | Q2NSH9 | 2-Hacid_dh | 9.40E-16 | - |
| VitaminB6 | *Sodalis* | Q2NSH9 | 2-Hacid_dh_C | 2.20E-30 | - |
| VitaminB6 | *Sodalis* | Q2NSH9 | DUF3410 | 1.40E-25 | - |
| VitaminB6 | *Ecoli* | P05459 | 2-Hacid_dh | 2.20E-18 | - |
| VitaminB6 | *Ecoli* | P05459 | 2-Hacid_dh_C | 8.30E-32 | - |
| VitaminB6 | *Ecoli* | P05459 | DUF3410 | 9.30E-24 | - |

| Pathway | Organism | Seq ID | Domain name | E-value | Predicted active site residues |
|---------|----------|--------|-------------|---------|-------------------------------|
| VitaminB6 | *Drosophilla* | Q9VAN0 | Aminotran_5 | 2.40E-50 | - |
| VitaminB6 | *Tribolium* | D6W9Q3 | Aminotran_5 | 6.70E-56 | - |
| VitaminB6 | *Anopheles* | Q5TRW7 | Aminotran_5 | 5.00E-44 | - |
| VitaminB6 | *Wigglesworthia* | H6Q4Q1 | Aminotran_5 | 4.40E-45 | - |
| VitaminB6 | *Sodalis* | Q2NUB0 | Aminotran_5 | 3.00E-56 | - |
| VitaminB6 | *Ecoli* | P23721 | Aminotran_5 | 4.30E-71 | - |
| VitaminB6 | *Bacillus* | P80862 | Aminotran_5 | 3.10E-44 | - |
| VitaminB6 | *Arabidopsis* | Q9SHP0 | Aminotran_5 | 2.00E-49 | - |
| VitaminB6 | *Arabidopsis* | Q96255 | Aminotran_5 | 1.90E-49 | - |
| VitaminB6 | *Yeast* | P0CX33 | Ribosomal_S30 | 9.60E-29 | - |
| VitaminB6 | *Human* | Q9Y617 | Aminotran_5 | 2.70E-67 | - |
| VitaminB6 | *Wigglesworthia* | H6Q499 | PdxA | 2.50E-95 | - |
| VitaminB6 | *Sodalis* | Q2NVX5 | PdxA | 5.60E-112 | - |
| VitaminB6 | *Ecoli* | P19624 | PdxA | 1.60E-123 | - |
| VitaminB6 | *Wigglesworthia* | H6Q5B7 | PdxJ | 7.00E-90 | [72,45,193] |
| VitaminB6 | *Sodalis* | Q2NS16 | PdxJ | 7.40E-102 | [72,45,193] |
| VitaminB6 | *Wolbachia* | Q3V8B3 | PdxJ | 3.30E-85 | [69,42,188] |
| VitaminB6 | *Ecoli* | P0A794 | PdxJ | 8.70E-106 | [72,45,193] |
| VitaminB6 | *Bacillus* | P37527 | SOR_SNZ | 8.30E-114 | - |
| VitaminB6 | *Arabidopsis* | O80448 | SOR_SNZ | 6.20E-115 | - |
| VitaminB6 | *Arabidopsis* | Q9ZNR6 | SOR_SNZ | 1.60E-77 | - |
| VitaminB6 | *Arabidopsis* | Q8L940 | SOR_SNZ | 9.00E-116 | - |
| VitaminB6 | *Sorghum* | C5X768 | SOR_SNZ | 5.80E-113 | - |
| VitaminB6 | *Yeast* | P43545 | SOR_SNZ | 1.00E-106 | - |
| VitaminB6 | *Yeast* | Q03148 | SOR_SNZ | 1.50E-112 | - |
| VitaminB6 | *Drosophilla* | Q7KUC2 | Phos_pyr_kin | 1.20E-15 | - |
| VitaminB6 | *Tribolium* | D6WEG5 | PfkB | 1.60E-18 | [223] |
| VitaminB6 | *Anopheles* | Q7Q6C1 | PfkB | 6.50E-18 | [231] |
| VitaminB6 | *Ecoli* | P77150 | Phos_pyr_kin | 2.20E-15 | - |
| VitaminB6 | *Ecoli* | P40191 | Phos_pyr_kin | 9.60E-26 | - |
| VitaminB6 | *Bacillus* | P39610 | Phos_pyr_kin | 1.30E-87 | - |
| VitaminB6 | *Arabidopsis* | Q8W1X2 | Phos_pyr_kin | 9.80E-16 | - |
| VitaminB6 | *Yeast* | P39988 | Phos_pyr_kin | 6.20E-11 | - |
| VitaminB6 | *Yeast* | P53727 | Phos_pyr_kin | 2.40E-10 | - |
| VitaminB6 | *Human* | O00764 | PfkB | 1.00E-33 | [235] |
| VitaminB6 | *Drosophilla* | Q9VWF0 | Put_Phosphatase | 4.20E-95 | - |
| VitaminB6 | *Tribolium* | D6WJQ0 | Put_Phosphatase | 1.00E-71 | - |
| VitaminB6 | *Anopheles* | Q7QCX1 | Put_Phosphatase | 6.30E-65 | - |
| VitaminB6 | *Human* | Q8TCD6 | Put_Phosphatase | 1.70E-103 | - |
| VitaminB6 | *Human* | Q96GD0 | Hydrolase_6 | 7.10E-26 | [25] |
| VitaminB6 | *Human* | Q96GD0 | Hydrolase_like | 3.50E-19 | - |
| VitaminB6 | *Drosophilla* | Q7KSW3 | Pyridox_oxidase | 1.90E-22 | - |
| VitaminB6 | *Drosophilla* | Q7KSW3 | PNPOx_C | 7.60E-22 | - |
| VitaminB6 | *Drosophilla* | Q8INR5 | Pyridox_oxidase | 3.00E-15 | - |
| VitaminB6 | *Anopheles* | Q7QK95 | Pyridox_oxidase | 1.40E-23 | - |
| VitaminB6 | *Anopheles* | Q7QK95 | PNPOx_C | 2.30E-22 | - |
| VitaminB6 | *Wigglesworthia* | H6Q4X8 | Pyridox_oxidase | 4.30E-17 | - |
| VitaminB6 | *Wigglesworthia* | H6Q4X8 | PNPOx_C | 3.80E-16 | - |
| VitaminB6 | *Sodalis* | Q2NT03 | Pyridox_oxidase | 3.00E-21 | - |
| VitaminB6 | *Sodalis* | Q2NT03 | PNPOx_C | 2.60E-19 | - |
| VitaminB6 | *Wolbachia* | Q73G09 | Pyridox_oxidase | 2.80E-23 | - |
| VitaminB6 | *Wolbachia* | Q73G09 | PNPOx_C | 1.90E-14 | - |
| VitaminB6 | *Ecoli* | P0AFI7 | Pyridox_oxidase | 6.00E-25 | - |
| VitaminB6 | *Ecoli* | P0AFI7 | PNPOx_C | 2.30E-19 | - |
| VitaminB6 | *Yeast* | P38075 | Pyridox_oxidase | 1.20E-12 | - |
| VitaminB6 | *Yeast* | P38075 | PNPOx_C | 4.60E-21 | - |
| VitaminB6 | *Human* | Q9NVS9 | Pyridox_oxidase | 6.30E-25 | - |
| VitaminB6 | *Human* | Q9NVS9 | PNPOx_C | 1.50E-19 | - |
| Niacin | *Sodalis* | Q2NS06 | FAD_binding_2 | 1.00E-100 | [244,263,290] |
| Niacin | *Sodalis* | Q2NS06 | Succ_DH_flav_C | 3.70E-19 | - |
| Niacin | *Ecoli* | P10902 | FAD_binding_2 | 1.20E-106 | [244,263,290] |
| Niacin | *Ecoli* | P10902 | Succ_DH_flav_C | 6.00E-19 | - |
| Niacin | *Bacillus* | P38032 | FAD_binding_2 | 3.80E-86 | [226,245,272] |
| Niacin | *Bacillus* | P38032 | Succ_DH_flav_C | 2.20E-15 | - |
| Niacin | *Arabidopsis* | Q94AY1 | FAD_binding_2 | 4.20E-88 | [313,341,368] |
| Niacin | *Arabidopsis* | Q94AY1 | Succ_DH_flav_C | 7.10E-20 | - |
| Niacin | *Sorghum* | C5XTX1 | FAD_binding_2 | 3.30E-90 | [315,343,370] |
| Niacin | *Sorghum* | C5XTX1 | Succ_DH_flav_C | 8.40E-20 | - |
| Niacin | *Wigglesworthia* | H6Q5F4 | NadA | 1.50E-100 | - |
| Niacin | *Sodalis* | Q2NUL1 | NadA | 3.70E-103 | - |
| Niacin | *Ecoli* | P11458 | NadA | 3.50E-105 | - |
| Niacin | *Bacillus* | Q9KWZ1 | NadA | 3.00E-107 | - |
| Niacin | *Arabidopsis* | Q9FGS4 | SufE | 9.30E-18 | - |

| Pathway | Organism | Seq ID | Domain name | E-value | Predicted active site residues |
|---------|----------|--------|-------------|---------|-------------------------------|
| Niacin | *Arabidopsis* | Q9FGS4 | NadA | 1.10E-29 | - |
| Niacin | *Sorghum* | C5YNE5 | NadA | 2.40E-32 | - |
| Niacin | *Ecoli* | P21369 | Isochorismatase | 1.80E-38 | [10,111,156] |
| Niacin | *Yeast* | P53184 | Isochorismatase | 4.50E-27 | [8,122,167] |
| Niacin | *Drosophilla* | Q9VQX4 | NAPRTase | 2.50E-31 | - |
| Niacin | *Anopheles* | Q7PJC3 | NAPRTase | 2.50E-10 | - |
| Niacin | *Anopheles* | Q7PJC3 | NAPRTase | 9.80E-19 | - |
| Niacin | *Sodalis* | Q2NU84 | NAPRTase | 2.70E-75 | - |
| Niacin | *Ecoli* | P18133 | NAPRTase | 4.70E-79 | - |
| Niacin | *Bacillus* | O32090 | NAPRTase | 2.30E-24 | - |
| Niacin | *Arabidopsis* | Q8RWM2 | NAPRTase | 2.20E-15 | - |
| Niacin | *Arabidopsis* | Q8RWM2 | NAPRTase | 1.20E-28 | - |
| Niacin | *Sorghum* | C5WTX0 | NAPRTase | 1.10E-11 | - |
| Niacin | *Sorghum* | C5WTX0 | NAPRTase | 5.30E-26 | - |
| Niacin | *Yeast* | P39683 | NAPRTase | 1.40E-75 | - |
| Niacin | *Human* | Q6XQN6 | NAPRTase | 8.20E-14 | - |
| Niacin | *Drosophilla* | Q7KS06 | CTP_transf_2 | 1.00E-32 | - |
| Niacin | *Tribolium* | D6X0P6 | CTP_transf_2 | 3.20E-34 | - |
| Niacin | *Anopheles* | Q7QG76 | CTP_transf_2 | 1.20E-32 | - |
| Niacin | *Wigglesworthia* | H6Q5E3 | CTP_transf_2 | 1.80E-25 | - |
| Niacin | *Sodalis* | Q2NUV0 | CTP_transf_2 | 1.60E-33 | - |
| Niacin | *Ecoli* | P0A752 | CTP_transf_2 | 3.50E-41 | - |
| Niacin | *Bacillus* | P54455 | CTP_transf_2 | 1.10E-40 | - |
| Niacin | *Arabidopsis* | F4K687 | CTP_transf_2 | 3.10E-30 | - |
| Niacin | *Sorghum* | C5XVU3 | CTP_transf_2 | 2.60E-31 | - |
| Niacin | *Yeast* | P53204 | CTP_transf_2 | 1.80E-35 | - |
| Niacin | *Yeast* | Q06178 | CTP_transf_2 | 3.10E-37 | - |
| Niacin | *Human* | Q9BZQ4 | CTP_transf_2 | 1.20E-32 | - |
| Niacin | *Human* | Q96T66 | CTP_transf_2 | 4.30E-47 | - |
| Niacin | *Human* | Q9HAN9 | CTP_transf_2 | 3.70E-34 | - |
| Niacin | *Drosophilla* | Q9VYA0 | CN_hydrolase | 4.50E-33 | [175] |
| Niacin | *Drosophilla* | Q9VYA0 | NAD_synthase | 3.50E-27 | - |
| Niacin | *Tribolium* | D1ZZT1 | CN_hydrolase | 5.00E-31 | [175] |
| Niacin | *Tribolium* | D1ZZT1 | NAD_synthase | 4.00E-27 | - |
| Niacin | *Anopheles* | Q7PS02 | CN_hydrolase | 3.70E-27 | [176] |
| Niacin | *Anopheles* | Q7PS02 | NAD_synthase | 6.10E-26 | - |
| Niacin | *Wigglesworthia* | H6Q4X3 | NAD_synthase | 2.60E-65 | - |
| Niacin | *Sodalis* | Q2NRT4 | NAD_synthase | 1.40E-71 | - |
| Niacin | *Ecoli* | P18843 | NAD_synthase | 9.40E-79 | - |
| Niacin | *Bacillus* | P08164 | NAD_synthase | 4.10E-82 | - |
| Niacin | *Arabidopsis* | Q9C723 | CN_hydrolase | 3.80E-22 | [174] |
| Niacin | *Arabidopsis* | Q9C723 | NAD_synthase | 3.10E-24 | - |
| Niacin | *Sorghum* | C5X4A1 | CN_hydrolase | 3.40E-21 | [174] |
| Niacin | *Sorghum* | C5X4A1 | NAD_synthase | 1.50E-27 | - |
| Niacin | *Yeast* | P38795 | CN_hydrolase | 1.50E-25 | [175] |
| Niacin | *Yeast* | P38795 | NAD_synthase | 5.00E-22 | - |
| Niacin | *Human* | Q6IA69 | CN_hydrolase | 1.70E-33 | [175] |
| Niacin | *Human* | Q6IA69 | NAD_synthase | 6.50E-27 | - |
| Niacin | *Drosophilla* | A1Z9F4 | NAD_kinase | 8.00E-62 | - |
| Niacin | *Anopheles* | Q7QHC1 | NAD_kinase | 1.50E-58 | - |
| Niacin | *Wigglesworthia* | H6Q5S1 | NAD_kinase | 6.70E-58 | - |
| Niacin | *Sodalis* | Q2NS01 | NAD_kinase | 8.10E-67 | - |
| Niacin | *Wolbachia* | Q73GR1 | NAD_kinase | 1.70E-29 | - |
| Niacin | *Ecoli* | P0A7B3 | NAD_kinase | 7.60E-73 | - |
| Niacin | *Bacillus* | O31612 | NAD_kinase | 1.20E-55 | - |
| Niacin | *Bacillus* | O34934 | NAD_kinase | 1.90E-51 | - |
| Niacin | *Arabidopsis* | Q56YN3 | NAD_kinase | 2.60E-61 | - |
| Niacin | *Sorghum* | C5XIJ6 | NAD_kinase | 3.60E-60 | - |
| Niacin | *Sorghum* | C5YXF6 | NAD_kinase | 1.50E-54 | - |
| Niacin | *Human* | O95544 | NAD_kinase | 8.80E-64 | - |
| Niacin | *Wigglesworthia* | H6Q4Z3 | QRPTase_N | 4.00E-27 | - |
| Niacin | *Wigglesworthia* | H6Q4Z3 | QRPTase_C | 1.80E-55 | - |
| Niacin | *Sodalis* | Q2NVT6 | QRPTase_N | 2.80E-25 | - |
| Niacin | *Sodalis* | Q2NVT6 | QRPTase_C | 2.10E-57 | - |
| Niacin | *Ecoli* | P30011 | QRPTase_N | 9.40E-27 | - |
| Niacin | *Ecoli* | P30011 | QRPTase_C | 4.60E-65 | - |
| Niacin | *Bacillus* | P39666 | QRPTase_N | 2.60E-25 | - |
| Niacin | *Bacillus* | P39666 | QRPTase_C | 3.00E-65 | - |
| Niacin | *Arabidopsis* | A8MRX1 | QRPTase_N | 8.30E-21 | - |
| Niacin | *Arabidopsis* | A8MRX1 | QRPTase_C | 2.50E-62 | - |
| Niacin | *Sorghum* | C5X7Q7 | QRPTase_N | 1.20E-23 | - |
| Niacin | *Sorghum* | C5X7Q7 | QRPTase_C | 2.30E-58 | - |
| Niacin | *Yeast* | P43619 | QRPTase_N | 5.20E-24 | - |

| Pathway | Organism | Seq ID | Domain name | E-value | Predicted active site residues |
|---|---|---|---|---|---|
| Niacin | *Yeast* | P43619 | QRPTase_C | 3.50E-60 | - |
| Niacin | *Human* | Q15274 | QRPTase_N | 3.20E-20 | - |
| Niacin | *Human* | Q15274 | QRPTase_C | 4.60E-59 | - |
| Coenzyme A | *Drosophilla* | Q9VYD5 | Aminotran_4 | 3.50E-18 | [279] |
| Coenzyme A | *Tribolium* | D6WAG1 | Aminotran_4 | 4.80E-20 | [253] |
| Coenzyme A | *Anopheles* | Q7QEI9 | Aminotran_4 | 5.80E-20 | [279] |
| Coenzyme A | *Sodalis* | Q2NQA5 | Aminotran_4 | 4.90E-41 | [160] |
| Coenzyme A | *Ecoli* | P0AB80 | Aminotran_4 | 1.40E-46 | [160] |
| Coenzyme A | *Bacillus* | O31461 | Aminotran_4 | 2.20E-31 | [197] |
| Coenzyme A | *Arabidopsis* | Q93Y32 | Aminotran_4 | 5.60E-31 | [231] |
| Coenzyme A | *Sorghum* | C5YVA1 | Aminotran_4 | 3.90E-40 | [181] |
| Coenzyme A | *Yeast* | P38891 | Aminotran_4 | 3.70E-29 | [219] |
| Coenzyme A | *Yeast* | P47176 | Aminotran_4 | 4.20E-30 | [202] |
| Coenzyme A | *Human* | P54687 | Aminotran_4 | 2.00E-23 | [222] |
| Coenzyme A | *Human* | O15382 | Aminotran_4 | 1.20E-22 | [229] |
| Coenzyme A | *Wigglesworthia* | H6Q5V2 | Pantoate_transf | 1.30E-102 | [181] |
| Coenzyme A | *Sodalis* | Q2NVR2 | Pantoate_transf | 1.10E-108 | [179] |
| Coenzyme A | *Ecoli* | P31057 | Pantoate_transf | 2.70E-107 | [181] |
| Coenzyme A | *Bacillus* | P52996 | Pantoate_transf | 5.50E-111 | [181] |
| Coenzyme A | *Arabidopsis* | O82357 | Pantoate_transf | 3.60E-104 | [222] |
| Coenzyme A | *Sorghum* | C5XKA7 | Pantoate_transf | 1.20E-104 | [255] |
| Coenzyme A | *Wigglesworthia* | H6Q5G7 | ApbA | 5.80E-27 | - |
| Coenzyme A | *Wigglesworthia* | H6Q5G7 | ApbA_C | 9.60E-28 | [176] |
| Coenzyme A | *Sodalis* | Q2NV89 | ApbA | 3.60E-27 | - |
| Coenzyme A | *Sodalis* | Q2NV89 | ApbA_C | 4.40E-33 | [176] |
| Coenzyme A | *Ecoli* | P0A9J4 | ApbA | 1.50E-28 | - |
| Coenzyme A | *Ecoli* | P0A9J4 | ApbA_C | 2.10E-39 | [176] |
| Coenzyme A | *Bacillus* | O31717 | ApbA | 2.70E-40 | - |
| Coenzyme A | *Bacillus* | O31717 | ApbA_C | 1.20E-24 | [184] |
| Coenzyme A | *Bacillus* | O34661 | ApbA | 2.70E-32 | - |
| Coenzyme A | *Bacillus* | O34661 | ApbA_C | 4.00E-29 | [179] |
| Coenzyme A | *Yeast* | P38787 | ApbA | 1.00E-35 | - |
| Coenzyme A | *Yeast* | P38787 | ApbA_C | 2.60E-24 | [224] |
| Coenzyme A | *Drosophilla* | Q9W374 | Fer4_20 | 5.90E-32 | - |
| Coenzyme A | *Drosophilla* | Q9W374 | NAD_binding_8 | 9.30E-08 | - |
| Coenzyme A | *Drosophilla* | Q9W374 | DHO_dh | 8.60E-37 | [666] |
| Coenzyme A | *Drosophilla* | Q9W374 | Fer4_21 | 1.30E-22 | - |
| Coenzyme A | *Tribolium* | D6WGA9 | Fer4_20 | 3.50E-31 | - |
| Coenzyme A | *Tribolium* | D6WGA9 | DHO_dh | 7.30E-33 | [662] |
| Coenzyme A | *Tribolium* | D6WGA9 | Fer4_21 | 6.80E-23 | - |
| Coenzyme A | *Anopheles* | Q7QHY0 | Fer4_20 | 2.00E-31 | - |
| Coenzyme A | *Anopheles* | Q7QHY0 | NAD_binding_8 | 6.10E-07 | - |
| Coenzyme A | *Anopheles* | Q7QHY0 | DHO_dh | 4.20E-33 | [669] |
| Coenzyme A | *Anopheles* | Q7QHY0 | Fer4_21 | 1.30E-22 | - |
| Coenzyme A | *Arabidopsis* | Q9LVI9 | DHO_dh | 2.70E-33 | [191] |
| Coenzyme A | *Sorghum* | C5XZ92 | DHO_dh | 9.20E-34 | [185] |
| Coenzyme A | *Human* | Q12882 | Fer4_20 | 3.90E-31 | - |
| Coenzyme A | *Human* | Q12882 | NAD_binding_8 | 6.50E-07 | - |
| Coenzyme A | *Human* | Q12882 | DHO_dh | 1.60E-35 | [671] |
| Coenzyme A | *Human* | Q12882 | Fer4_21 | 5.90E-23 | - |
| Coenzyme A | *Drosophilla* | Q8IPQ2 | Amidohydro_1 | 4.20E-21 | - |
| Coenzyme A | *Anopheles* | F5HKZ6 | Amidohydro_5 | 8.50E-13 | - |
| Coenzyme A | *Ecoli* | Q46806 | Amidohydro_4 | 8.90E-28 | - |
| Coenzyme A | *Arabidopsis* | Q9FMP3 | Amidohydro_4 | 7.50E-16 | - |
| Coenzyme A | *Sorghum* | C5XMK0 | Amidohydro_4 | 6.00E-19 | - |
| Coenzyme A | *Human* | Q14117 | Amidohydro_1 | 9.30E-25 | - |
| Coenzyme A | *Drosophilla* | Q9VI04 | CN_hydrolase | 2.50E-39 | [234,197,120] |
| Coenzyme A | *Tribolium* | D2A4C0 | CN_hydrolase | 7.70E-37 | [232,195,118] |
| Coenzyme A | *Anopheles* | Q7Q0P4 | CN_hydrolase | 3.60E-38 | [234,197,120] |
| Coenzyme A | *Arabidopsis* | Q8H183 | CN_hydrolase | 4.20E-36 | [249,212,137] |
| Coenzyme A | *Sorghum* | C5X8L4 | CN_hydrolase | 3.70E-37 | [255,218,143] |
| Coenzyme A | *Human* | Q9UBR1 | CN_hydrolase | 2.80E-40 | [233,196,119] |
| Coenzyme A | *Sodalis* | Q2NVR4 | Asp_decarbox | 1.70E-50 | [25,58] |
| Coenzyme A | *Ecoli* | P0A790 | Asp_decarbox | 3.50E-49 | [25,58] |
| Coenzyme A | *Bacillus* | P52999 | Asp_decarbox | 3.40E-54 | [25,58] |
| Coenzyme A | *Wigglesworthia* | H6Q5V3 | Pantoate_ligase | 2.90E-88 | [39] |
| Coenzyme A | *Sodalis* | Q2NVR3 | Pantoate_ligase | 3.60E-102 | [37] |
| Coenzyme A | *Ecoli* | P31663 | Pantoate_ligase | 8.00E-124 | [37] |
| Coenzyme A | *Bacillus* | P52998 | Pantoate_ligase | 3.30E-127 | [37] |
| Coenzyme A | *Arabidopsis* | Q9FKB3 | Pantoate_ligase | 2.40E-91 | [39] |
| Coenzyme A | *Sorghum* | C5WS23 | Pantoate_ligase | 3.20E-86 | [44] |
| Coenzyme A | *Yeast* | P40459 | Pantoate_ligase | 2.70E-125 | [40] |
| Coenzyme A | *Drosophilla* | D8FT20 | Fumble | 3.20E-130 | [69] |

94

| Pathway | Organism | Seq ID | Domain name | E-value | Predicted active site residues |
|---------|----------|--------|-------------|---------|-------------------------------|
| Coenzyme A | *Drosophilla* | Q9VMU2 | DUF89 | 2.70E-54 | - |
| Coenzyme A | *Tribolium* | D6WCF9 | Fumble | 2.20E-137 | [137] |
| Coenzyme A | *Tribolium* | D6WMP7 | DUF89 | 6.90E-42 | - |
| Coenzyme A | *Anopheles* | Q7Q0Y3 | DUF89 | 1.90E-47 | - |
| Coenzyme A | *Anopheles* | Q7PVC2 | Fumble | 1.80E-138 | [146] |
| Coenzyme A | *Wigglesworthia* | H6Q4M3 | PRK | 8.70E-11 | - |
| Coenzyme A | *Sodalis* | Q2NWS4 | PRK | 1.50E-13 | - |
| Coenzyme A | *Ecoli* | P0A6I3 | PRK | 5.60E-13 | - |
| Coenzyme A | *Bacillus* | P37564 | Pan_kinase | 1.10E-70 | - |
| Coenzyme A | *Bacillus* | P54556 | PRK | 1.60E-14 | - |
| Coenzyme A | *Arabidopsis* | Q8L5Y9 | Fumble | 1.20E-146 | [201] |
| Coenzyme A | *Arabidopsis* | Q8L5Y9 | DUF89 | 1.20E-44 | - |
| Coenzyme A | *Sorghum* | C5X682 | Fumble | 2.70E-138 | [205] |
| Coenzyme A | *Sorghum* | C5X682 | DUF89 | 4.80E-44 | - |
| Coenzyme A | *Yeast* | Q04430 | Fumble | 1.30E-132 | [105] |
| Coenzyme A | *Human* | Q8TE04 | Fumble | 3.10E-146 | [363] |
| Coenzyme A | *Human* | Q9NVE7 | Fumble | 6.70E-143 | - |
| Coenzyme A | *Human* | Q9NVE7 | DUF89 | 1.00E-57 | - |
| Coenzyme A | *Human* | Q9H999 | Fumble | 4.70E-144 | [138] |
| Coenzyme A | *Human* | Q9BZ23 | Fumble | 3.00E-139 | [338] |
| Coenzyme A | *Drosophilla* | Q7KN99 | DFP | 4.00E-06 | - |
| Coenzyme A | *Drosophilla* | Q7KN99 | DFP | 2.90E-11 | - |
| Coenzyme A | *Tribolium* | D6X537 | DFP | 4.00E-11 | - |
| Coenzyme A | *Anopheles* | Q7QAC3 | DFP | 5.90E-06 | - |
| Coenzyme A | *Anopheles* | Q7QAC3 | DFP | 2.70E-11 | - |
| Coenzyme A | *Wigglesworthia* | H6Q4R3 | Flavoprotein | 4.90E-34 | [75] |
| Coenzyme A | *Wigglesworthia* | H6Q4R3 | DFP | 3.40E-66 | - |
| Coenzyme A | *Sodalis* | Q2NQU1 | Flavoprotein | 8.00E-32 | [76] |
| Coenzyme A | *Sodalis* | Q2NQU1 | DFP | 1.80E-69 | - |
| Coenzyme A | *Ecoli* | P0ABQ0 | Flavoprotein | 5.40E-31 | [75] |
| Coenzyme A | *Ecoli* | P0ABQ0 | DFP | 1.10E-72 | - |
| Coenzyme A | *Bacillus* | O35033 | Flavoprotein | 4.80E-35 | [74] |
| Coenzyme A | *Bacillus* | O35033 | DFP | 3.60E-72 | - |
| Coenzyme A | *Arabidopsis* | Q8GXR5 | DFP | 9.50E-13 | - |
| Coenzyme A | *Arabidopsis* | Q9LZM3 | DFP | 3.00E-14 | - |
| Coenzyme A | *Sorghum* | C5XVP2 | DFP | 1.10E-07 | - |
| Coenzyme A | *Sorghum* | C5XVP2 | DFP | 1.30E-11 | - |
| Coenzyme A | *Sorghum* | C5YHW1 | DFP | 5.10E-08 | - |
| Coenzyme A | *Sorghum* | C5YHW1 | DFP | 2.20E-13 | - |
| Coenzyme A | *Yeast* | P40506 | DFP | 4.90E-06 | - |
| Coenzyme A | *Human* | Q9HAB8 | DFP | 1.20E-06 | - |
| Coenzyme A | *Human* | Q9HAB8 | DFP | 1.00E-11 | - |
| Coenzyme A | *Drosophilla* | Q8MKK3 | Flavoprotein | 9.30E-41 | [80] |
| Coenzyme A | *Tribolium* | D2CFX2 | Flavoprotein | 2.10E-44 | [76] |
| Coenzyme A | *Anopheles* | Q7PZN2 | Flavoprotein | 2.90E-45 | [79] |
| Coenzyme A | *Arabidopsis* | P94063 | Flavoprotein | 2.40E-42 | [82] |
| Coenzyme A | *Arabidopsis* | Q9SWE5 | Flavoprotein | 1.40E-42 | [90] |
| Coenzyme A | *Sorghum* | C5XLV1 | Flavoprotein | 4.20E-41 | [88] |
| Coenzyme A | *Sorghum* | C5Z606 | Flavoprotein | 1.60E-38 | [91] |
| Coenzyme A | *Sorghum* | C5Z607 | Flavoprotein | 1.10E-42 | [91] |
| Coenzyme A | *Yeast* | P36076 | Flavoprotein | 9.20E-39 | [391] |
| Coenzyme A | *Human* | Q96CD2 | Flavoprotein | 1.30E-44 | [88] |
| Coenzyme A | *Drosophilla* | Q9VRP4 | CTP_transf_2 | 3.50E-10 | - |
| Coenzyme A | *Drosophilla* | Q9VRP4 | CoaE | 4.90E-44 | - |
| Coenzyme A | *Anopheles* | Q7Q774 | CTP_transf_2 | 2.20E-10 | - |
| Coenzyme A | *Anopheles* | Q7Q774 | CoaE | 8.30E-40 | - |
| Coenzyme A | *Wigglesworthia* | H6Q528 | CTP_transf_2 | 3.30E-19 | - |
| Coenzyme A | *Sodalis* | Q2NQU5 | CTP_transf_2 | 1.20E-20 | - |
| Coenzyme A | *Ecoli* | P0A6I6 | CTP_transf_2 | 8.50E-25 | - |
| Coenzyme A | *Bacillus* | O34797 | CTP_transf_2 | 4.50E-23 | - |
| Coenzyme A | *Arabidopsis* | Q9ZPV8 | CTP_transf_2 | 3.30E-11 | - |
| Coenzyme A | *Sorghum* | C5XB13 | CTP_transf_2 | 2.50E-09 | - |
| Coenzyme A | *Sorghum* | C5XTW9 | CTP_transf_2 | 4.50E-13 | - |
| Coenzyme A | *Yeast* | P53332 | CTP_transf_2 | 2.60E-20 | - |
| Coenzyme A | *Human* | Q13057 | CTP_transf_2 | 4.80E-09 | - |
| Coenzyme A | *Human* | Q13057 | CoaE | 4.70E-40 | - |
| Coenzyme A | *Wigglesworthia* | H6Q5C1 | CoaE | 1.80E-44 | - |
| Coenzyme A | *Sodalis* | Q2NVT9 | CoaE | 1.90E-72 | - |
| Coenzyme A | *Ecoli* | P0A6I9 | CoaE | 1.10E-77 | - |
| Coenzyme A | *Bacillus* | O34932 | CoaE | 2.50E-65 | - |
| Coenzyme A | *Arabidopsis* | Q9ZQH0 | CoaE | 1.50E-61 | - |
| Coenzyme A | *Sorghum* | C5XKI1 | CoaE | 1.10E-61 | - |
| Coenzyme A | *Sorghum* | C5YLK2 | CoaE | 3.40E-63 | - |

| Pathway | Organism | Seq ID | Domain name | E-value | Predicted active site residues |
|---|---|---|---|---|---|
| Coenzyme A | *Yeast* | Q03941 | CoaE | 2.20E-70 | - |
| THF | *Drosophilla* | P48596 | GTP_cyclohydroI | 3.40E-72 | - |
| THF | *Tribolium* | D2CFW6 | GTP_cyclohydroI | 1.90E-73 | - |
| THF | *Anopheles* | A7UU97 | GTP_cyclohydroI | 1.50E-74 | - |
| THF | *Wigglesworthia* | H6Q500 | GTP_cyclohydroI | 1.10E-59 | - |
| THF | *Sodalis* | Q2NUE3 | GTP_cyclohydroI | 7.60E-63 | - |
| THF | *Ecoli* | P0A6T5 | GTP_cyclohydroI | 1.50E-63 | - |
| THF | *Bacillus* | P19465 | GTP_cyclohydroI | 5.60E-81 | - |
| THF | *Arabidopsis* | F4JED5 | GTP_cyclohydroI | 4.60E-35 | - |
| THF | *Arabidopsis* | F4JED5 | GTP_cyclohydroI | 6.50E-39 | - |
| THF | *Yeast* | P51601 | GTP_cyclohydroI | 7.60E-75 | - |
| THF | *Human* | P30793 | GTP_cyclohydroI | 2.00E-75 | - |
| THF | *Drosophilla* | Q9VRM8 | Alk_phosphatase | 6.20E-128 | [128,128] |
| THF | *Tribolium* | D2A3C2 | Alk_phosphatase | 4.00E-120 | [127,127] |
| THF | *Anopheles* | Q7PY02 | Alk_phosphatase | 1.20E-134 | [153,153] |
| THF | *Ecoli* | P00634 | Alk_phosphatase | 1.20E-134 | [124,124] |
| THF | *Ecoli* | P00634 | Alk_phosphatase | 2.30E-09 | - |
| THF | *Bacillus* | P42251 | PhoD | 9.80E-165 | - |
| THF | *Bacillus* | P19405 | Alk_phosphatase | 7.70E-153 | [101,101] |
| THF | *Bacillus* | P19406 | Alk_phosphatase | 2.20E-151 | [108,108] |
| THF | *Arabidopsis* | F4K1J1 | PhoD | 7.90E-26 | - |
| THF | *Sorghum* | C5XWF9 | PhoD | 9.90E-25 | - |
| THF | *Yeast* | P11491 | Alk_phosphatase | 3.10E-158 | [123,123] |
| THF | *Human* | P09923 | Alk_phosphatase | 5.20E-185 | [111,111] |
| THF | *Human* | P05186 | Alk_phosphatase | 1.60E-180 | [110,110] |
| THF | *Human* | P05187 | Alk_phosphatase | 1.00E-188 | [114,114] |
| THF | *Human* | P10696 | Alk_phosphatase | 2.60E-187 | [111,111] |
| THF | *Wigglesworthia* | H6Q5S8 | FolB | 1.00E-24 | - |
| THF | *Sodalis* | Q2NWE4 | FolB | 8.80E-31 | - |
| THF | *Ecoli* | P0AC16 | FolB | 2.20E-32 | - |
| THF | *Bacillus* | P28823 | FolB | 1.70E-31 | - |
| THF | *Arabidopsis* | Q9SF23 | FolB | 1.00E-26 | - |
| THF | *Arabidopsis* | F4IYU3 | FolB | 1.40E-22 | - |
| THF | *Arabidopsis* | Q9FM54 | FolB | 2.90E-23 | - |
| THF | *Sorghum* | C5YNA8 | FolB | 3.30E-29 | - |
| THF | *Yeast* | P53848 | FolB | 7.40E-17 | - |
| THF | *Yeast* | P53848 | FolB | 3.10E-16 | - |
| THF | *Yeast* | P53848 | HPPK | 1.80E-38 | - |
| THF | *Yeast* | P53848 | Pterin_bind | 5.10E-42 | - |
| THF | *Wigglesworthia* | H6Q551 | HPPK | 7.00E-42 | - |
| THF | *Sodalis* | Q2NVR1 | HPPK | 2.60E-43 | - |
| THF | *Wolbachia* | Q73GQ3 | Pterin_bind | 5.90E-29 | - |
| THF | *Ecoli* | P26281 | HPPK | 2.80E-44 | - |
| THF | *Bacillus* | P29252 | HPPK | 4.20E-48 | - |
| THF | *Arabidopsis* | F4JPH1 | Pterin_bind | 3.50E-77 | - |
| THF | *Sorghum* | C5X2E7 | HPPK | 1.40E-39 | - |
| THF | *Sorghum* | C5X2E7 | Pterin_bind | 1.30E-72 | - |
| THF | *Sorghum* | C5XIR9 | HPPK | 3.90E-39 | - |
| THF | *Sorghum* | C5XIR9 | Pterin_bind | 1.40E-72 | - |
| THF | *Wigglesworthia* | H6Q5R8 | Anth_synt_I_N | 1.90E-19 | - |
| THF | *Wigglesworthia* | H6Q5R8 | Chorismate_bind | 5.70E-74 | - |
| THF | *Sodalis* | Q2NTC1 | Anth_synt_I_N | 7.50E-27 | - |
| THF | *Sodalis* | Q2NTC1 | Chorismate_bind | 1.20E-88 | - |
| THF | *Sodalis* | Q2NQJ8 | GATase | 1.20E-57 | [174,176,81] |
| THF | *Ecoli* | P05041 | Anth_synt_I_N | 2.60E-28 | - |
| THF | *Ecoli* | P05041 | Chorismate_bind | 2.90E-90 | - |
| THF | *Ecoli* | P00903 | GATase | 2.20E-61 | [168,170,79] |
| THF | *Bacillus* | P28820 | Anth_synt_I_N | 7.60E-32 | - |
| THF | *Bacillus* | P28820 | Chorismate_bind | 1.80E-95 | - |
| THF | *Bacillus* | P28819 | GATase | 2.90E-60 | [168,170,79] |
| THF | *Arabidopsis* | Q8LPN3 | GATase | 1.40E-28 | [172] |
| THF | *Arabidopsis* | Q8LPN3 | GATase | 2.70E-07 | - |
| THF | *Arabidopsis* | Q8LPN3 | Anth_synt_I_N | 5.20E-20 | - |
| THF | *Arabidopsis* | Q8LPN3 | Chorismate_bind | 4.80E-87 | - |
| THF | *Sorghum* | C5Z8W2 | GATase | 4.70E-09 | - |
| THF | *Sorghum* | C5Z8W2 | GATase | 2.00E-09 | - |
| THF | *Sorghum* | C5Z8W2 | Anth_synt_I_N | 4.80E-19 | - |
| THF | *Sorghum* | C5Z8W2 | Chorismate_bind | 4.60E-88 | - |
| THF | *Yeast* | P37254 | GATase | 1.50E-32 | [207,209,112] |
| THF | *Yeast* | P37254 | Anth_synt_I_N | 4.10E-13 | - |
| THF | *Yeast* | P37254 | Chorismate_bind | 2.40E-63 | - |
| THF | *Wigglesworthia* | H6Q5Q3 | Aminotran_4 | 7.60E-34 | [141] |
| THF | *Sodalis* | Q2NU37 | Aminotran_4 | 1.10E-43 | [140] |

| Pathway | Organism | Seq ID | Domain name | E-value | Predicted active site residues |
|---------|----------|--------|-------------|---------|-------------------------------|
| THF | *Ecoli* | P28305 | Aminotran_4 | 2.60E-42 | [140] |
| THF | *Bacillus* | P28821 | Aminotran_4 | 3.50E-44 | [146] |
| THF | *Wigglesworthia* | H6Q587 | Pterin_bind | 1.60E-62 | - |
| THF | *Sodalis* | Q2NW28 | Pterin_bind | 2.40E-68 | - |
| THF | *Ecoli* | P0AC13 | Pterin_bind | 2.10E-76 | - |
| THF | *Bacillus* | P28822 | Pterin_bind | 4.90E-75 | - |
| THF | *Wigglesworthia* | H6Q506 | Mur_ligase_M | 4.70E-10 | - |
| THF | *Sodalis* | Q2NSI4 | Mur_ligase_M | 4.40E-13 | - |
| THF | *Wolbachia* | Q73GA7 | Mur_ligase_M | 3.80E-11 | - |
| THF | *Ecoli* | P08192 | Mur_ligase_M | 9.70E-14 | - |
| THF | *Ecoli* | P08192 | Mur_ligase_C | 7.80E-07 | - |
| THF | *Bacillus* | Q05865 | Mur_ligase_M | 6.30E-16 | - |
| THF | *Bacillus* | Q05865 | Mur_ligase_C | 2.30E-11 | - |
| THF | *Arabidopsis* | F4J2K2 | Mur_ligase_M | 6.40E-06 | - |
| THF | *Arabidopsis* | Q8W035 | Mur_ligase_M | 1.20E-06 | - |
| THF | *Arabidopsis* | F4JYE9 | Mur_ligase_M | 2.90E-11 | - |
| THF | *Yeast* | Q12676 | Mur_ligase_M | 2.00E-06 | - |
| THF | *Drosophilla* | P17719 | DHFR_1 | 5.50E-37 | - |
| THF | *Anopheles* | Q7Q0L5 | DHFR_1 | 9.50E-40 | - |
| THF | *Wigglesworthia* | H6Q4A2 | DHFR_1 | 2.70E-52 | - |
| THF | *Sodalis* | Q2NVX9 | DHFR_1 | 2.90E-56 | - |
| THF | *Wolbachia* | Q73GQ2 | DHFR_1 | 9.30E-37 | - |
| THF | *Ecoli* | P0ABQ4 | DHFR_1 | 7.50E-59 | - |
| THF | *Bacillus* | P11045 | DHFR_1 | 2.20E-65 | - |
| THF | *Arabidopsis* | Q05762 | DHFR_1 | 1.80E-37 | - |
| THF | *Arabidopsis* | Q05762 | Thymidylat_synt | 6.50E-121 | [401] |
| THF | *Arabidopsis* | Q05763 | DHFR_1 | 2.40E-36 | - |
| THF | *Arabidopsis* | Q05763 | Thymidylat_synt | 1.30E-122 | [447] |
| THF | *Sorghum* | C5Y2E9 | DHFR_1 | 1.00E-33 | - |
| THF | *Sorghum* | C5Y2E9 | Thymidylat_synt | 1.90E-125 | [403] |
| THF | *Yeast* | P07807 | DHFR_1 | 3.20E-30 | - |
| THF | *Human* | P00374 | DHFR_1 | 2.20E-34 | - |
| THF | *Human* | Q86XF0 | DHFR_1 | 2.10E-31 | - |
| Biotin | *Wigglesworthia* | H6Q5U6 | Methyltransf_11 | 7.40E-25 | - |
| Biotin | *Sodalis* | Q2NUJ5 | Methyltransf_11 | 3.70E-25 | - |
| Biotin | *Ecoli* | P12999 | Methyltransf_11 | 2.60E-23 | - |
| Biotin | *Wigglesworthia* | H6Q5Z4 | ketoacyl-synt | 3.80E-40 | [161,161] |
| Biotin | *Wigglesworthia* | H6Q5Z4 | Ketoacyl-synt_C | 1.90E-26 | - |
| Biotin | *Sodalis* | Q2NSH7 | ketoacyl-synt | 1.10E-40 | [161,161] |
| Biotin | *Sodalis* | Q2NSH7 | Ketoacyl-synt_C | 4.70E-29 | - |
| Biotin | *Ecoli* | P0A953 | ketoacyl-synt | 9.10E-43 | [163,163] |
| Biotin | *Ecoli* | P0A953 | Ketoacyl-synt_C | 2.30E-29 | - |
| Biotin | *Drosophilla* | Q9VNF5 | ketoacyl-synt | 3.90E-58 | [187,187] |
| Biotin | *Drosophilla* | Q9VNF5 | Ketoacyl-synt_C | 1.10E-32 | - |
| Biotin | *Tribolium* | D6X1D1 | ketoacyl-synt | 5.50E-57 | [167,167] |
| Biotin | *Tribolium* | D6X1D1 | Ketoacyl-synt_C | 9.30E-32 | - |
| Biotin | *Anopheles* | Q7QCU6 | ketoacyl-synt | 1.80E-57 | [203,203] |
| Biotin | *Anopheles* | Q7QCU6 | Ketoacyl-synt_C | 3.80E-34 | - |
| Biotin | *Sodalis* | Q2NU38 | ketoacyl-synt | 2.80E-56 | [164,164] |
| Biotin | *Sodalis* | Q2NU38 | Ketoacyl-synt_C | 4.80E-40 | - |
| Biotin | *Wolbachia* | Q73FX9 | ketoacyl-synt | 5.00E-56 | [174,174] |
| Biotin | *Wolbachia* | Q73FX9 | Ketoacyl-synt_C | 4.30E-38 | - |
| Biotin | *Ecoli* | P0AAI5 | ketoacyl-synt | 1.00E-59 | [164,164] |
| Biotin | *Ecoli* | P0AAI5 | Ketoacyl-synt_C | 3.10E-39 | - |
| Biotin | *Bacillus* | O34340 | ketoacyl-synt | 1.00E-62 | [164,164] |
| Biotin | *Bacillus* | O34340 | Ketoacyl-synt_C | 1.40E-40 | - |
| Biotin | *Arabidopsis* | Q8L3X9 | ketoacyl-synt | 7.40E-54 | [209,209] |
| Biotin | *Arabidopsis* | Q8L3X9 | Ketoacyl-synt_C | 1.50E-35 | - |
| Biotin | *Sorghum* | C5Z5Z9 | ketoacyl-synt | 1.70E-55 | [212,212] |
| Biotin | *Sorghum* | C5Z5Z9 | Ketoacyl-synt_C | 1.80E-32 | - |
| Biotin | *Yeast* | P39525 | ketoacyl-synt | 3.70E-56 | [187,187] |
| Biotin | *Yeast* | P39525 | Ketoacyl-synt_C | 9.40E-33 | - |
| Biotin | *Human* | Q9NWU1 | ketoacyl-synt | 2.50E-55 | [209,209] |
| Biotin | *Human* | Q9NWU1 | Ketoacyl-synt_C | 7.50E-36 | - |
| Biotin | *Wigglesworthia* | H6Q5Q1 | adh_short | 6.40E-33 | [151] |
| Biotin | *Sodalis* | Q2NU40 | adh_short | 6.40E-37 | [151] |
| Biotin | *Wolbachia* | Q73HB7 | adh_short | 3.70E-36 | [152] |
| Biotin | *Ecoli* | P0AEK2 | adh_short | 9.50E-40 | [151] |
| Biotin | *Bacillus* | P51831 | adh_short | 8.90E-45 | [154] |
| Biotin | *Bacillus* | O31767 | adh_short | 2.70E-26 | [150] |
| Biotin | *Bacillus* | O34308 | adh_short_C2 | 1.10E-21 | - |
| Biotin | *Arabidopsis* | P33207 | adh_short | 4.20E-43 | [226] |
| Biotin | *Arabidopsis* | Q9SQR4 | adh_short | 1.20E-30 | [178] |

| Pathway | Organism | Seq ID | Domain name | E-value | Predicted active site residues |
|---|---|---|---|---|---|
| Biotin | *Arabidopsis* | Q9SQR2 | adh_short_C2 | 2.70E-32 | [183,176] |
| Biotin | *Arabidopsis* | Q9SVQ9 | adh_short_C2 | 1.30E-30 | [174,167] |
| Biotin | *Arabidopsis* | F4JWJ4 | adh_short | 1.90E-30 | [168] |
| Biotin | *Sorghum* | C5WZI4 | adh_short | 5.40E-36 | [166] |
| Biotin | *Sorghum* | C5WZI6 | adh_short | 1.90E-36 | [170] |
| Biotin | *Sorghum* | C5WZI7 | adh_short | 2.50E-35 | [170] |
| Biotin | *Sorghum* | C5WWL1 | adh_short | 7.90E-22 | [183] |
| Biotin | *Sorghum* | C5XAC7 | adh_short_C2 | 3.60E-29 | [187,180] |
| Biotin | *Sorghum* | C5X9U6 | adh_short | 1.00E-28 | [182] |
| Biotin | *Sorghum* | C5XSJ4 | adh_short | 3.20E-43 | [223] |
| Biotin | *Sorghum* | C5XUE9 | adh_short_C2 | 2.10E-33 | [206,199] |
| Biotin | *Sorghum* | C5YE75 | adh_short | 1.40E-42 | [223] |
| Biotin | *Wigglesworthia* | H6Q4T3 | FabA | 3.10E-41 | [54] |
| Biotin | *Sodalis* | Q2NRL8 | FabA | 4.40E-42 | [54] |
| Biotin | *Wolbachia* | P61455 | FabA | 2.30E-35 | [49] |
| Biotin | *Ecoli* | P0A6Q6 | FabA | 1.10E-42 | [54] |
| Biotin | *Bacillus* | P94584 | FabA | 2.10E-45 | [48] |
| Biotin | *Arabidopsis* | Q9SIE3 | FabA | 2.00E-37 | [122] |
| Biotin | *Arabidopsis* | Q9LX13 | FabA | 1.30E-37 | [121] |
| Biotin | *Sorghum* | C5YIY2 | FabA | 4.20E-38 | [127] |
| Biotin | *Sorghum* | C5YYP0 | FabA | 4.50E-37 | [119] |
| Biotin | *Wigglesworthia* | H6Q4V4 | adh_short_C2 | 3.10E-76 | [164,157] |
| Biotin | *Sodalis* | Q2NST7 | adh_short_C2 | 1.10E-80 | [163,156] |
| Biotin | *Wolbachia* | Q73IR6 | adh_short_C2 | 1.10E-78 | [164,157] |
| Biotin | *Ecoli* | P0AEK4 | adh_short_C2 | 2.40E-79 | [163,156] |
| Biotin | *Bacillus* | P54616 | adh_short_C2 | 6.70E-82 | [165,158] |
| Biotin | *Arabidopsis* | Q9SLA8 | adh_short_C2 | 2.20E-101 | [282,274] |
| Biotin | *Wigglesworthia* | H6Q4F0 | Abhydrolase_6 | 3.10E-26 | [82,236,207] |
| Biotin | *Sodalis* | Q2NQH6 | Abhydrolase_6 | 1.20E-29 | [82,235,207] |
| Biotin | *Ecoli* | P13001 | Abhydrolase_6 | 6.80E-34 | [82,235,207] |
| Biotin | *Wigglesworthia* | H6Q5U7 | Aminotran_1_2 | 2.40E-43 | [256] |
| Biotin | *Sodalis* | Q2NUJ6 | Aminotran_1_2 | 2.10E-61 | [238] |
| Biotin | *Ecoli* | P12998 | Aminotran_1_2 | 1.30E-69 | [236] |
| Biotin | *Bacillus* | P53556 | Aminotran_1_2 | 6.20E-64 | [237] |
| Biotin | *Arabidopsis* | Q2QKD2 | Aminotran_1_2 | 9.40E-49 | [319] |
| Biotin | *Sorghum* | C5WSC4 | Aminotran_1_2 | 2.40E-45 | [301] |
| Biotin | *Wigglesworthia* | H6Q5U9 | Aminotran_3 | 2.40E-98 | - |
| Biotin | *Sodalis* | Q2NUJ8 | Aminotran_3 | 1.70E-108 | - |
| Biotin | *Ecoli* | P12995 | Aminotran_3 | 8.60E-114 | - |
| Biotin | *Bacillus* | P53555 | Aminotran_3 | 2.10E-107 | - |
| Biotin | *Arabidopsis* | B0F481 | AAA_26 | 9.30E-44 | - |
| Biotin | *Arabidopsis* | B0F481 | Aminotran_3 | 7.10E-44 | - |
| Biotin | *Sorghum* | C5WTT1 | AAA_26 | 1.90E-40 | - |
| Biotin | *Sorghum* | C5WTT1 | Aminotran_3 | 1.40E-43 | - |
| Biotin | *Yeast* | P50277 | Aminotran_3 | 5.10E-84 | - |
| Biotin | *Wigglesworthia* | H6Q5U5 | AAA_26 | 1.20E-34 | - |
| Biotin | *Sodalis* | Q2NUJ4 | AAA_26 | 1.80E-35 | - |
| Biotin | *Sodalis* | Q2NSY4 | AAA_26 | 1.60E-23 | - |
| Biotin | *Sodalis* | Q2NSY4 | MFS_1 | 1.60E-14 | - |
| Biotin | *Ecoli* | P13000 | AAA_26 | 4.70E-42 | - |
| Biotin | *Ecoli* | P0A6E9 | AAA_26 | 2.90E-30 | - |
| Biotin | *Bacillus* | P53558 | AAA_26 | 1.00E-49 | - |
| Biotin | *Yeast* | P53630 | AAA_26 | 1.40E-49 | - |

## Table S3: Physicochemical properties of orthologs

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---|---|---|---|---|---|---|
| Thiamine | *Glossina* | 2.8.1.7 | TMP002270 | 467 | 8.36 | 52129.56 |
| Thiamine | *Glossina* | 2.8.1.7 | TMP003624 | 293 | 5.87 | 31752.83 |
| Thiamine | *Glossina* | 2.8.1.7 | TMP002651 | 176 | 8.97 | 19627.6 |
| Thiamine | *Glossina* | 2.8.1.7 | TMP006982 | 148 | 6.63 | 16110.55 |
| Thiamine | *Glossina* | 2.8.1.7 | TMPEC2.8.1.7-648456 | 314 | 5.71 | 34074.21 |
| Thiamine | *Glossina* | 2.8.1.7 | TMPEC2.8.1.7-641678 | 134 | 9.55 | 14765.82 |
| Thiamine | *Glossina* | 2.8.1.7 | TMPEC2.8.1.7-652413 | 218 | 6.24 | 23377.14 |
| Thiamine | *Glossina* | 2.8.1.7 | TMPEC2.8.1.7-641518 | 257 | 6.43 | 27888.49 |
| Thiamine | *Drosophilla* | 2.8.1.7 | Q9VKD3 | 462 | 8.4 | 51074.18 |
| Thiamine | *Tribolium* | 2.8.1.7 | D6WLY0 | 445 | 8.5 | 49273.5 |
| Thiamine | *Anopheles* | 2.8.1.7 | Q7PW45 | 462 | 8.32 | 49288.61 |
| Thiamine | *Wigglesworthia* | 2.8.1.7 | H6Q4V9 | 410 | 9.62 | 46519.71 |
| Thiamine | *Wigglesworthia* | 2.8.1.7 | H6Q534 | 381 | 9.48 | 42470.38 |
| Thiamine | *Sodalis* | 2.8.1.7 | Q2NT15 | 407 | 5.67 | 44886.42 |

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---|---|---|---|---|---|---|
| Thiamine | *Sodalis* | 2.8.1.7 | Q2NS31 | 399 | 6.73 | 42889.12 |
| Thiamine | *Wolbachia* | 2.8.1.7 | Q73H67 | 378 | 5.67 | 40913.05 |
| Thiamine | *Wolbachia* | 2.8.1.7 | Q73GF8 | 415 | 6.55 | 46234.02 |
| Thiamine | *Escherichia* | 2.8.1.7; 4.4.1.16 | P77444 | 406 | 5.89 | 44433.76 |
| Thiamine | *Escherichia* | 2.8.1.7 | P0A6B7 | 404 | 5.94 | 45089.5 |
| Thiamine | *Bacillus* | 2.8.1.7 | O34599 | 379 | 5.55 | 41191.56 |
| Thiamine | *Bacillus* | 2.8.1.7 | P38033 | 395 | 6.41 | 43785.97 |
| Thiamine | *Bacillus* | 2.8.1.7 | O34874 | 381 | 6.33 | 41487.55 |
| Thiamine | *Bacillus* | 2.8.1.7 | O32164 | 406 | 5.32 | 44921.99 |
| Thiamine | *Arabidopsis* | 2.8.1.7; 4.4.1.16 | Q93WX6 | 463 | 5.98 | 46761.08 |
| Thiamine | *Arabidopsis* | 2.8.1.7 | O49543 | 453 | 6.52 | 50295.75 |
| Thiamine | *Sorghum* | 2.8.1.7 | C5Y0K0 | 375 | 5.89 | 41293.01 |
| Thiamine | *Saccharomyces* | 2.8.1.7 | P25374 | 497 | 6.63 | 50723.63 |
| Thiamine | *Human* | 2.8.1.7 | Q9Y697 | 457 | 8.54 | 50195.67 |
| Thiamine | *Glossina* | ThiC | TMPThiC-641665 | 253 | 5.36 | 28340.58 |
| Thiamine | *Wigglesworthia* | ThiC | H6Q4N0 | 623 | 8.95 | 71576.86 |
| Thiamine | *Escherichia* | ThiC | P30136 | 631 | 5.7 | 70850.36 |
| Thiamine | *Bacillus* | ThiC | P45740 | 590 | 5.44 | 65931.72 |
| Thiamine | *Arabidopsis* | ThiC | O82392 | 644 | 5.83 | 68066.07 |
| Thiamine | *Sorghum* | ThiC | C5WPN8 | 640 | 5.9 | 71291 |
| Thiamine | *Glossina* | ThiD | TMPThiD-641665 | 81 | 6.63 | 8868.05 |
| Thiamine | *Wigglesworthia* | ThiD | H6Q4Z9 | 268 | 9.51 | 29395.4 |
| Thiamine | *Escherichia* | ThiD | P76422 | 266 | 5.73 | 28633.61 |
| Thiamine | *Bacillus* | ThiD | O31620 | 271 | 5.74 | 29124.03 |
| Thiamine | *Arabidopsis* | ThiD | Q5M731 | 522 | 5.76 | 51867.44 |
| Thiamine | *Sorghum* | ThiD | C5YTC8 | 547 | 7.5 | 57782.46 |
| Thiamine | *Saccharomyces* | ThiD | Q08224 | 551 | 5.81 | 61269.35 |
| Thiamine | *Saccharomyces* | ThiD | Q08975 | 551 | 5.63 | 61334.3 |
| Thiamine | *Escherichia* | ThiF | P30138 | 251 | 4.69 | 26969.78 |
| Thiamine | *Bacillus* | ThiF | O31619 | 336 | 5.72 | 36399.71 |
| Thiamine | *Wigglesworthia* | ThiS | H6Q4N2 | 66 | 8.01 | 7439.76 |
| Thiamine | *Escherichia* | ThiS | O32583 | 66 | 4.37 | 7311.29 |
| Thiamine | *Bacillus* | ThiS | O31617 | 66 | 5.59 | 7625.79 |
| Thiamine | *Wigglesworthia* | ThiH | H6Q4N4 | 372 | 9.05 | 43387.34 |
| Thiamine | *Sodalis* | ThiH | Q2NWR4 | 278 | 9.33 | 32065.97 |
| Thiamine | *Escherichia* | ThiH | P30140 | 377 | 6.57 | 43320.13 |
| Thiamine | *Bacillus* | ThiO | O31616 | 369 | 5.92 | 40936.83 |
| Thiamine | *Wigglesworthia* | ThiG | H6Q4N3 | 254 | 9.07 | 27828.52 |
| Thiamine | *Escherichia* | ThiG | P30139 | 256 | 5.36 | 26896.1 |
| Thiamine | *Bacillus* | ThiG | O31618 | 256 | 4.91 | 27022.23 |
| Thiamine | *Wigglesworthia* | 2.5.1.3 | H6Q4N1 | 212 | 9.74 | 24365.79 |
| Thiamine | *Escherichia* | 2.5.1.3 | P30137 | 211 | 5.51 | 23015.28 |
| Thiamine | *Bacillus* | 2.5.1.3 | P39594 | 222 | 5.24 | 23680.96 |
| Thiamine | *Arabidopsis* | 2.5.1.3; 2.7.1.49 | Q5M731 | 522 | 5.76 | 51867.44 |
| Thiamine | *Sorghum* | 2.5.1.3 | C5YTC8 | 547 | 7.5 | 57782.46 |
| Thiamine | *Saccharomyces* | 2.5.1.3; 2.7.1.50 | P41835 | 540 | 5.67 | 58058.71 |
| Thiamine | *Wigglesworthia* | 2.7.4.16 | H6Q5T5 | 288 | 9.71 | 32668.31 |
| Thiamine | *Sodalis* | 2.7.4.16 | Q2NV96 | 328 | 4.95 | 36055.83 |
| Thiamine | *Escherichia* | 2.7.4.16 | P0AGG0 | 325 | 4.48 | 35070.85 |
| Thiamine | *Bacillus* | 2.7.4.16 | O05514 | 325 | 5.19 | 35883.99 |
| Thiamine | *Glossina* | 2.7.6.2 | TMP011237 | 291 | 8.08 | 33322.16 |
| Thiamine | *Drosophilla* | 2.7.6.2 | Q8T4A5 | 345 | 9.08 | 38383.85 |
| Thiamine | *Tribolium* | 2.7.6.2 | 664370/XP_975470.2 | 322 | 6.07 | 29464.7 |
| Thiamine | *Anopheles* | 2.7.6.2 | Q7QD56 | 271 | 5.49 | 29879.9 |
| Thiamine | *Bacillus* | 2.7.6.2 | O34664 | 214 | 5.34 | 24098.59 |
| Thiamine | *Arabidopsis* | 2.7.6.2 | F4IV16 | 267 | 5.7 | 30214.43 |
| Thiamine | *Sorghum* | 2.7.6.2 | C5XJY7 | 267 | 6.89 | 30442.84 |
| Thiamine | *Sorghum* | 2.7.6.2 | C5XH23 | 263 | 5.1 | 28943.79 |
| Thiamine | *Sorghum* | 2.7.6.2 | C5YWW8 | 277 | 4.99 | 30085.04 |
| Thiamine | *Saccharomyces* | 2.7.6.2 | P35202 | 319 | 5.92 | 36616.09 |
| Thiamine | *Human* | 2.7.6.2 | Q9H3S4 | 243 | 5.03 | 27265.28 |
| Riboflavin | *Wigglesworthia* | 4.1.99.12 | H6Q580 | 214 | 9.06 | 23546.54 |
| Riboflavin | *Sodalis* | 4.1.99.12 | Q2NWD7 | 217 | 5.15 | 23535.77 |
| Riboflavin | *Wolbachia* | 4.1.99.12 | Q73HB4 | 217 | 5.98 | 23909.6 |
| Riboflavin | *Escherichia* | 4.1.99.12 | P0A7J0 | 217 | 4.9 | 23353.47 |
| Riboflavin | *Bacillus* | 4.1.99.12; 3.5.4.25 | P17620 | 398 | 5.64 | 44121.41 |
| Riboflavin | *Arabidopsis* | 4.1.99.12 | Q6NLQ7 | 476 | 5.8 | 52199.79 |
| Riboflavin | *Arabidopsis* | 4.1.99.12 | F4KJA1 | 543 | 5.8 | 60176.6 |
| Riboflavin | *Arabidopsis* | 4.1.99.12; 3.5.4.25 | P47924 | 543 | 5.1 | 53093.25 |
| Riboflavin | *Sorghum* | 4.1.99.12 | C5XVM3 | 554 | 5.92 | 59885.25 |
| Riboflavin | *Sorghum* | 4.1.99.12 | C5YJR6 | 547 | 5.54 | 59397.41 |
| Riboflavin | *Sorghum* | 4.1.99.12 | C5YZA3 | 603 | 5.92 | 65823.04 |
| Riboflavin | *Saccharomyces* | 4.1.99.12 | Q99258 | 208 | 5.46 | 22567.65 |

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---|---|---|---|---|---|---|
| Riboflavin | *Glossina* | 3.5.4.25 | TMP003195 | 379 | 5.66 | 41912.38 |
| Riboflavin | *Glossina* | 3.5.4.25 | TMPEC3.5.4.25-648457 | 301 | 6.46 | 34169.06 |
| Riboflavin | *Wigglesworthia* | 3.5.4.25 | H6Q4V2 | 199 | 8.75 | 22313.77 |
| Riboflavin | *Sodalis* | 3.5.4.25 | Q2NT40 | 199 | 5.9 | 21896.16 |
| Riboflavin | *Wolbachia* | 3.5.4.25 | Q73IY9 | 360 | 5.97 | 41129.81 |
| Riboflavin | *Escherichia* | 3.5.4.25 | P0A7I7 | 196 | 5.6 | 21835.98 |
| Riboflavin | *Bacillus* | 4.1.99.12; 3.5.4.25 | P17620 | 398 | 5.64 | 44121.41 |
| Riboflavin | *Arabidopsis* | 3.5.4.25 | Q6NLQ7 | 476 | 5.8 | 52199.79 |
| Riboflavin | *Arabidopsis* | 3.5.4.25 | F4KJA1 | 543 | 5.8 | 60176.6 |
| Riboflavin | *Arabidopsis* | 4.1.99.12; 3.5.4.25 | P47924 | 543 | 5.1 | 53093.25 |
| Riboflavin | *Sorghum* | 3.5.4.25 | C5XVM3 | 554 | 5.92 | 59885.25 |
| Riboflavin | *Sorghum* | 3.5.4.25 | C5YJR6 | 547 | 5.54 | 59397.41 |
| Riboflavin | *Sorghum* | 3.5.4.25 | C5YZA3 | 603 | 5.92 | 65823.04 |
| Riboflavin | *Saccharomyces* | 3.5.4.25 | P38066 | 345 | 5.9 | 38331.95 |
| Riboflavin | *Glossina* | 3.5.4.26 | TMPEC1.1.1.193-648456 | 325 | 8.31 | 35652.32 |
| Riboflavin | *Wigglesworthia* | 3.5.4.26 | H6Q5T9 | 380 | 9.84 | 42681.84 |
| Riboflavin | *Sodalis* | 3.5.4.26 | Q2NV99 | 375 | 6.76 | 40511.45 |
| Riboflavin | *Wolbachia* | 3.5.4.26 | Q73H62 | 360 | 8.58 | 39424.8 |
| Riboflavin | *Escherichia* | 3.5.4.26; 1.1.1.193 | P25539 | 367 | 7.22 | 40338.31 |
| Riboflavin | *Bacillus* | 3.5.4.26; 1.1.1.193 | P17618 | 361 | 6.09 | 39305.31 |
| Riboflavin | *Arabidopsis* | 3.5.4.26 | Q8GWP5 | 426 | 6.04 | 46673.67 |
| Riboflavin | *Sorghum* | 3.5.4.26 | C5X102 | 396 | 7.14 | 42288.43 |
| Riboflavin | *Wigglesworthia* | 2.5.1.78 | H6Q5T8 | 154 | 9.3 | 17108.2 |
| Riboflavin | *Sodalis* | 2.5.1.78 | Q2NV98 | 156 | 5.35 | 16056.5 |
| Riboflavin | *Wolbachia* | 2.5.1.78 | P61729 | 142 | 6.96 | 15467.93 |
| Riboflavin | *Escherichia* | 2.5.1.78 | P61714 | 156 | 5.15 | 16156.51 |
| Riboflavin | *Bacillus* | 2.5.1.78 | P11998 | 154 | 5.35 | 16286.62 |
| Riboflavin | *Arabidopsis* | 2.5.1.78 | O80575 | 227 | 6.22 | 16542.98 |
| Riboflavin | *Sorghum* | 2.5.1.78 | C5Y9F9 | 215 | 8.93 | 22270.55 |
| Riboflavin | *Saccharomyces* | 2.5.1.78 | P50861 | 169 | 6.06 | 18555.52 |
| Riboflavin | *Glossina* | 2.5.1.9 | TMPEC2.5.1.9-638809 | 198 | 5.83 | 22287.44 |
| Riboflavin | *Wigglesworthia* | 2.5.1.9 | H6Q4Y2 | 204 | 8.83 | 22654.47 |
| Riboflavin | *Sodalis* | 2.5.1.9 | Q2NT11 | 205 | 6.16 | 22361.99 |
| Riboflavin | *Wolbachia* | 2.5.1.9 | Q73IM7 | 198 | 5.82 | 22380.53 |
| Riboflavin | *Escherichia* | 2.5.1.9 | P0AFU8 | 213 | 5.64 | 23444.9 |
| Riboflavin | *Bacillus* | 2.5.1.9 | P16440 | 215 | 5.87 | 23480.96 |
| Riboflavin | *Arabidopsis* | 2.5.1.9 | Q9SKU8 | 271 | 7.62 | 29639.25 |
| Riboflavin | *Saccharomyces* | 2.5.1.9 | P38145 | 238 | 5.04 | 26195.78 |
| Riboflavin | *Glossina* | 2.7.1.26 | TMP011242 | 165 | 5.2 | 18617.98 |
| Riboflavin | *Glossina* | 2.7.1.26 | TMPEC2.7.1.26-642265 | 315 | 8.55 | 36223.76 |
| Riboflavin | *Drosophilla* | 2.7.1.26 | O76206 | 153 | 5.92 | 16985.5 |
| Riboflavin | *Tribolium* | 2.7.1.26 | 660247/XP_971589.1 | 148 | 5.41 | 16937.28 |
| Riboflavin | *Wigglesworthia* | 2.7.1.26 | H6Q522 | 312 | 10.07 | 35777.4 |
| Riboflavin | *Sodalis* | 2.7.1.26 | Q2NVY7 | 312 | 9.68 | 34360.41 |
| Riboflavin | *Wolbachia* | 2.7.1.26 | Q73H17 | 310 | 8.09 | 35559.04 |
| Riboflavin | *Escherichia* | 2.7.1.26; 2.7.7.2 | P0AG40 | 313 | 9.34 | 34734.27 |
| Riboflavin | *Bacillus* | 2.7.1.26; 2.7.7.2 | P54575 | 316 | 8.26 | 35661.91 |
| Riboflavin | *Arabidopsis* | 2.7.1.26 | Q84MD8 | 379 | 5.97 | 42110.43 |
| Riboflavin | *Sorghum* | 2.7.1.26 | C5X020 | 396 | 5.72 | 43377.52 |
| Riboflavin | *Saccharomyces* | 2.7.1.26 | Q03778 | 218 | 5.03 | 22366.41 |
| Riboflavin | *Human* | 2.7.1.26 | Q969G6 | 155 | 7.85 | 17623.17 |
| Riboflavin | *Glossina* | 2.7.7.2 | TMP002859a | 424 | 9.34 | 49225.29 |
| Riboflavin | *Glossina* | 2.7.7.2 | TMP010445 | 247 | 6.33 | 28852.25 |
| Riboflavin | *Drosophilla* | 2.7.7.2 | Q9VJY1 | 254 | 6.02 | 29722.84 |
| Riboflavin | *Drosophilla* | 2.7.7.2 | Q8IR76 | 294 | 5.71 | 33699.19 |
| Riboflavin | *Tribolium* | 2.7.7.2 | D6WDE0 | 455 | 6.26 | 52908.47 |
| Riboflavin | *Anopheles* | 2.7.7.2 | F5HKN2 | 188 | 5.55 | 21534.47 |
| Riboflavin | *Wigglesworthia* | 2.7.7.2 | H6Q522 | 312 | 10.07 | 35777.4 |
| Riboflavin | *Sodalis* | 2.7.7.2 | Q2NVY7 | 312 | 9.68 | 34360.41 |
| Riboflavin | *Wolbachia* | 2.7.7.2 | Q73H17 | 310 | 8.09 | 35559.04 |
| Riboflavin | *Escherichia* | 2.7.1.26; 2.7.7.2 | P0AG40 | 313 | 9.34 | 34734.27 |
| Riboflavin | *Bacillus* | 2.7.1.26; 2.7.7.2 | P54575 | 316 | 8.26 | 35661.91 |
| Riboflavin | *Saccharomyces* | 2.7.7.2 | P38913 | 306 | 5.16 | 35546.16 |
| Riboflavin | *Human* | 2.7.7.2 | Q8NFF5 | 587 | 6.49 | 65265.57 |
| Riboflavin | *Glossina* | 1.5.1.30 | TMP005236 | 202 | 6.31 | 22469.84 |
| Riboflavin | *Human* | 1.5.1.30; 1.3.1.24 | P30043 | 206 | 7.31 | 21988.16 |
| Pyridoxine | *Escherichia* | 1.2.1.72 | P0A9B6 | 339 | 6.26 | 37168.2 |
| Pyridoxine | *Glossina* | 1.1.1.290 | TMP003196 | 318 | 5.34 | 35162.5 |
| Pyridoxine | *Glossina* | 1.1.1.290 | TMP005771 | 461 | 6.31 | 49747.92 |
| Pyridoxine | *Wigglesworthia* | 1.1.1.290 | H6Q509 | 381 | 9.76 | 43516.58 |
| Pyridoxine | *Sodalis* | 1.1.1.290 | Q2NSH9 | 377 | 5.37 | 40428.29 |
| Pyridoxine | *Escherichia* | 1.1.1.290 | P05459 | 378 | 6.23 | 41367.65 |
| Pyridoxine | *Glossina* | 2.6.1.52 | TMP008908 | 364 | 8.73 | 40227.45 |

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---|---|---|---|---|---|---|
| Pyridoxine | *Drosophilla* | 2.6.1.52 | Q9VAN0 | 364 | 8.52 | 39540.36 |
| Pyridoxine | *Tribolium* | 2.6.1.52 | D6W9Q3 | 368 | 6.55 | 40495.31 |
| Pyridoxine | *Anopheles* | 2.6.1.52 | Q5TRW7 | 364 | 7.65 | 40126.34 |
| Pyridoxine | *Wigglesworthia* | 2.6.1.52 | H6Q4Q1 | 362 | 9.76 | 41619.69 |
| Pyridoxine | *Sodalis* | 2.6.1.52 | Q2NUB0 | 361 | 5.54 | 39898.45 |
| Pyridoxine | *Escherichia* | 2.6.1.52 | P23721 | 362 | 5.37 | 39652.12 |
| Pyridoxine | *Bacillus* | 2.6.1.52 | P80862 | 359 | 5.62 | 40135.64 |
| Pyridoxine | *Arabidopsis* | 2.6.1.52 | Q9SHP0 | 422 | 8.26 | 46633.52 |
| Pyridoxine | *Arabidopsis* | 2.6.1.52 | Q96255 | 430 | 8.25 | 47359.32 |
| Pyridoxine | *Saccharomyces* | ^2.6.1.52 | P33330 | 63 | 6.08 | 43415.57 |
| Pyridoxine | *Human* | 2.6.1.52 | Q9Y617 | 370 | 7.56 | 40422.68 |
| Pyridoxine | *Glossina* | 1.1.1.262 | TMP003625 | 573 | 5.81 | 62239.73 |
| Pyridoxine | *Wigglesworthia* | 1.1.1.262 | H6Q499 | 328 | 9.21 | 36364 |
| Pyridoxine | *Sodalis* | 1.1.1.262 | Q2NVX5 | 330 | 5.63 | 35035.26 |
| Pyridoxine | *Escherichia* | 1.1.1.262 | P19624 | 329 | 5.87 | 35113.78 |
| Pyridoxine | *Wigglesworthia* | 2.6.99.2 | H6Q5B7 | 244 | 9.22 | 27299.95 |
| Pyridoxine | *Sodalis* | 2.6.99.2 | Q2NS16 | 243 | 5.61 | 26336.13 |
| Pyridoxine | *Wolbachia* | 2.6.99.2 | Q3V8B3 | 235 | 6.31 | 26347.36 |
| Pyridoxine | *Escherichia* | 2.6.99.2 | P0A794 | 243 | 5.61 | 26253.09 |
| Pyridoxine | *Bacillus* | 4.3.3.6 | P37527 | 294 | 5.26 | 31480.36 |
| Pyridoxine | *Arabidopsis* | 4.3.3.6 | O80448 | 309 | 5.73 | 32861.92 |
| Pyridoxine | *Arabidopsis* | 4.3.3.6 | Q9ZNR6 | 314 | 5.45 | 33835.54 |
| Pyridoxine | *Arabidopsis* | 4.3.3.6 | Q8L940 | 309 | 5.77 | 33216.31 |
| Pyridoxine | *Sorghum* | 4.3.3.6 | C5X768 | 317 | 6.33 | 33621.61 |
| Pyridoxine | *Saccharomyces* | 4.3.3.6 | P43545 | 298 | 5.4 | 32019.02 |
| Pyridoxine | *Saccharomyces* | 4.3.3.6 | Q03148 | 297 | 5.4 | 31816.86 |
| Pyridoxine | *Glossina* | 2.7.1.35 | TMP005733 | 304 | 5.71 | 33836.8 |
| Pyridoxine | *Glossina* | 2.7.1.35 | TMPEC2.7.1.35-639530 | 47 | 6.76 | 5328.18 |
| Pyridoxine | *Drosophilla* | 2.7.1.35 | Q7KUC2 | 304 | 6.46 | 33428.42 |
| Pyridoxine | *Tribolium* | 2.7.1.35 | D6WEG5 | 294 | 6.66 | 32131.05 |
| Pyridoxine | *Anopheles* | 2.7.1.35 | Q7Q6C1 | 309 | 6.46 | 34124.2 |
| Pyridoxine | *Escherichia* | 2.7.1.35 | P77150 | 287 | 6.04 | 31322.2 |
| Pyridoxine | *Escherichia* | 2.7.1.35 | P40191 | 283 | 5.14 | 30847.4 |
| Pyridoxine | *Bacillus* | 2.7.1.35 | P39610 | 271 | 5.09 | 29017.26 |
| Pyridoxine | *Arabidopsis* | 2.7.1.35 | Q8W1X2 | 309 | 5.55 | 34043.06 |
| Pyridoxine | *Saccharomyces* | 2.7.1.35 | P39988 | 312 | 6.11 | 35559.19 |
| Pyridoxine | *Saccharomyces* | 2.7.1.35 | P53727 | 317 | 6.32 | 35366.8 |
| Pyridoxine | *Human* | 2.7.1.35 | O00764 | 312 | 5.75 | 35102.3 |
| Pyridoxine | *Glossina* | 3.1.3.- | TMP011872 | 275 | 8.53 | 31928.82 |
| Pyridoxine | *Drosophilla* | 3.1.3.- | Q9VWF0 | 306 | 7.64 | 34621.44 |
| Pyridoxine | *Tribolium* | 3.1.3.74 | D6WJQ0 | 245 | 5.98 | 27714.86 |
| Pyridoxine | *Anopheles* | 3.1.3.74 | Q7QCX1 | 250 | 5.91 | 28279.16 |
| Pyridoxine | *Human* | 3.1.3.74 | Q8TCD6 | 241 | 6.3 | 27768.91 |
| Pyridoxine | *Human* | 3.1.3.3; 3.1.3.74 | Q96GD0 | 296 | 6.11 | 31698.15 |
| Pyridoxine | *Glossina* | 1.4.3.5 | TMP007245 | 254 | 8.82 | 29217.53 |
| Pyridoxine | *Drosophilla* | 1.4.3.5 | Q7KSW3 | 237 | 6.36 | 26787.35 |
| Pyridoxine | *Drosophilla* | 1.4.3.5 | Q8INR5 | 257 | 9.12 | 29294.2 |
| Pyridoxine | *Tribolium* | 1.4.3.5 | 660386/XP_971715.1 | 226 | 7.68 | 26079.43 |
| Pyridoxine | *Anopheles* | 1.4.3.5 | Q7QK95 | 241 | 6.88 | 27844.83 |
| Pyridoxine | *Wigglesworthia* | 1.4.3.5 | H6Q4X8 | 216 | 10.51 | 26161.89 |
| Pyridoxine | *Sodalis* | 1.4.3.5 | Q2NT03 | 216 | 9.66 | 25029.63 |
| Pyridoxine | *Wolbachia* | 1.4.3.5 | Q73G09 | 216 | 8.71 | 25349.29 |
| Pyridoxine | *Escherichia* | 1.4.3.5 | P0AFI7 | 218 | 9.18 | 25413.96 |
| Pyridoxine | *Saccharomyces* | 1.4.3.5 | P38075 | 228 | 6.94 | 26908.26 |
| Pyridoxine | *Human* | 1.4.3.5 | Q9NVS9 | 261 | 6.61 | 29988 |
| Niacin | *Sodalis* | 1.4.3.16 | Q2NS06 | 533 | 6.16 | 59283.35 |
| Niacin | *Escherichia* | 1.4.3.16 | P10902 | 540 | 5.89 | 60337.41 |
| Niacin | *Bacillus* | 1.4.3.16 | P38032 | 531 | 6.43 | 58239.33 |
| Niacin | *Arabidopsis* | 1.4.3.16 | Q94AY1 | 651 | 6.67 | 71411.71 |
| Niacin | *Sorghum* | 1.4.3.16 | C5XTX1 | 654 | 6.33 | 72032.89 |
| Niacin | *Wigglesworthia* | NadA(2.5.1.72) | H6Q5F4 | 344 | 9.18 | 38546.26 |
| Niacin | *Sodalis* | NadA(2.5.1.72) | Q2NUL1 | 348 | 5.71 | 37817.57 |
| Niacin | *Escherichia* | NadA(2.5.1.72) | P11458 | 347 | 5.19 | 38240.8 |
| Niacin | *Bacillus* | NadA(2.5.1.72) | Q9KWZ1 | 368 | 5.76 | 41492.63 |
| Niacin | *Arabidopsis* | NadA(2.5.1.72) | Q9FGS4 | 718 | 6.42 | 78933.76 |
| Niacin | *Sorghum* | NadA(2.5.1.72) | C5YNE5 | 580 | 6.02 | 63322.3 |
| Niacin | *Glossina* | 3.5.1.19 | TMP008696 | 357 | 4.74 | 40183.25 |
| Niacin | *Escherichia* | 3.5.1.-; 3.5.1.19 | P21369 | 213 | 4.6 | 23362.05 |
| Niacin | *Saccharomyces* | 3.5.1.19 | P53184 | 216 | 5.81 | 24993.31 |
| Niacin | *Glossina* | 2.4.2.11 | TMP011160 | 480 | 9.01 | 53256.65 |
| Niacin | *Glossina* | 2.4.2.11 | TMP003848 | 204 | 5.11 | 23006.01 |
| Niacin | *Drosophilla* | 2.4.2.11 | Q9VQX4 | 555 | 6.32 | 62165.2 |
| Niacin | *Tribolium* | 2.4.2.11 | 657856/XP_001814986.1 | 550 | 6.81 | 61626.09 |

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---|---|---|---|---|---|---|
| Niacin | *Anopheles* | 2.4.2.11 | Q7PJC3 | 519 | 5.72 | 58111.45 |
| Niacin | *Sodalis* | 2.4.2.11 | Q2NU84 | 402 | 8.93 | 46361.33 |
| Niacin | *Escherichia* | 2.4.2.11 | P18133 | 400 | 6.2 | 45766.09 |
| Niacin | *Bacillus* | 2.4.2.11 | O32090 | 490 | 5.25 | 56179.69 |
| Niacin | *Arabidopsis* | 2.4.2.11 | Q84WV8 | 557 | 5.86 | 62273.28 |
| Niacin | *Arabidopsis* | 2.4.2.11 | Q8RWM2 | 559 | 5.91 | 62420.6 |
| Niacin | *Sorghum* | 2.4.2.11 | C5WTX0 | 557 | 6.55 | 62125.16 |
| Niacin | *Sorghum* | 2.4.2.11 | C5YGH6 | 536 | 6.38 | 59840.79 |
| Niacin | *Saccharomyces* | 2.4.2.11 | P39683 | 429 | 6.44 | 49018.6 |
| Niacin | *Human* | 2.4.2.11 | Q6XQN6 | 538 | 5.51 | 57578.23 |
| Niacin | *Glossina* | 2.7.7.1/2.7.7.18 | TMP004826 | 359 | 8.69 | 40801.81 |
| Niacin | *Drosophilla* | 2.7.7.1 | Q7KS06 | 297 | 9 | 33455.29 |
| Niacin | *Tribolium* | 2.7.7.1/2.7.7.18 | D6X0P6 | 400 | 8.26 | 45350.68 |
| Niacin | *Anopheles* | 2.7.7.1/2.7.7.18 | Q7QG76 | 246 | 8.39 | 27836.85 |
| Niacin | *Wigglesworthia* | 2.7.7.18 | H6Q5E3 | 210 | 9.56 | 24279.72 |
| Niacin | *Sodalis* | 2.7.7.18 | Q2NUV0 | 218 | 7.74 | 24505.11 |
| Niacin | *Escherichia* | 2.7.7.18 | P0A752 | 213 | 5.46 | 24527.95 |
| Niacin | *Bacillus* | 2.7.7.18 | P54455 | 189 | 5.28 | 22156.5 |
| Niacin | *Arabidopsis* | 2.7.7.1/2.7.7.18 | F4K687 | 238 | 5.77 | 26922.89 |
| Niacin | *Sorghum* | 2.7.7.1/2.7.7.18 | C5XVU3 | 251 | 5.75 | 28006.25 |
| Niacin | *Saccharomyces* | 2.7.7.1 | P53204 | 395 | 5.57 | 44909.07 |
| Niacin | *Saccharomyces* | 2.7.7.1 | Q06178 | 401 | 6.4 | 45859.05 |
| Niacin | *Human* | 2.7.7.1; 2.7.7.18 | Q9BZQ4 | 307 | 6.59 | 34438.63 |
| Niacin | *Human* | 2.7.7.18; 2.7.7.1 | Q96T66 | 252 | 9.28 | 28321.67 |
| Niacin | *Human* | 2.7.7.1; 2.7.7.18 | Q9HAN9 | 279 | 8.98 | 31932.46 |
| Niacin | *Glossina* | 6.3.5.1 | TMP010364 | 865 | 6.76 | 97065.02 |
| Niacin | *Glossina* | 6.3.5.1 | TMPEC6.3.1.5-641666 | 180 | 3.75 | 21131.73 |
| Niacin | *Drosophilla* | 6.3.5.1 | Q9VYA0 | 787 | 6.23 | 87614.79 |
| Niacin | *Tribolium* | 6.3.5.1 | D1ZZT1 | 724 | 7.71 | 81438.39 |
| Niacin | *Anopheles* | 6.3.5.1 | Q7PS02 | 825 | 6.51 | 89358.16 |
| Niacin | *Wigglesworthia* | 6.3.1.5 | H6Q4X3 | 267 | 8.91 | 30241.89 |
| Niacin | *Sodalis* | 6.3.1.5 | Q2NRT4 | 274 | 5.25 | 30207.4 |
| Niacin | *Escherichia* | 6.3.1.5 | P18843 | 275 | 5.41 | 30636.83 |
| Niacin | *Bacillus* | 6.3.1.5 | P08164 | 272 | 5.07 | 30264.02 |
| Niacin | *Arabidopsis* | 6.3.5.1 | Q9C723 | 725 | 5.61 | 80900.44 |
| Niacin | *Sorghum* | 6.3.5.1 | C5X4A1 | 732 | 5.99 | 82114.41 |
| Niacin | *Saccharomyces* | 6.3.5.1 | P38795 | 714 | 6.11 | 80685.69 |
| Niacin | *Human* | 6.3.5.1 | Q6IA69 | 706 | 6.02 | 79284.73 |
| Niacin | *Glossina* | 2.7.1.23 | TMP012832 | 544 | 6.05 | 61252.41 |
| Niacin | *Glossina* | 2.7.1.23 | TMPEC2.7.1.23-650872 | 173 | 6.06 | 19195 |
| Niacin | *Glossina* | 2.7.1.23 | TMPEC2.7.1.23-641851 | 112 | 6.1 | 11994.79 |
| Niacin | *Drosophilla* | 2.7.1.23 | A1Z9F4 | 490 | 6.65 | 54793.36 |
| Niacin | *Tribolium* | 2.7.1.23 | 663340/XP_974485.1 | 497 | 7.61 | 55549.89 |
| Niacin | *Anopheles* | 2.7.1.23 | Q7QHC1 | 535 | 6.25 | 59588.73 |
| Niacin | *Wigglesworthia* | 2.7.1.23 | H6Q5S1 | 293 | 9.83 | 33370.76 |
| Niacin | *Sodalis* | 2.7.1.23 | Q2NS01 | 292 | 5.69 | 32060.62 |
| Niacin | *Wolbachia* | 2.7.1.23 | Q73GR1 | 264 | 6.8 | 30059.3 |
| Niacin | *Escherichia* | 2.7.1.23 | P0A7B3 | 292 | 6.3 | 32566.38 |
| Niacin | *Bacillus* | 2.7.1.23 | O31612 | 266 | 6 | 30012.34 |
| Niacin | *Bacillus* | 2.7.1.23 | O34934 | 267 | 6.45 | 30253.4 |
| Niacin | *Arabidopsis* | 2.7.1.23; 2.7.1.86 | Q56YN3 | 524 | 6.04 | 58244.99 |
| Niacin | *Sorghum* | 2.7.1.23 | C5XIJ6 | 462 | 6.64 | 51378.59 |
| Niacin | *Sorghum* | 2.7.1.23 | C5YXF6 | 498 | 6.37 | 54805.42 |
| Niacin | *Human* | 2.7.1.23 | O95544 | 446 | 6.03 | 49228.21 |
| Niacin | *Wigglesworthia* | 2.4.2.19 | H6Q4Z3 | 288 | 9.27 | 32568.95 |
| Niacin | *Sodalis* | 2.4.2.19 | Q2NVT6 | 296 | 5.2 | 31571.13 |
| Niacin | *Escherichia* | 2.4.2.19 | P30011 | 297 | 5.07 | 32630.87 |
| Niacin | *Bacillus* | 2.4.2.19 | P39666 | 289 | 5.31 | 31392.93 |
| Niacin | *Arabidopsis* | 2.4.2.19 | A8MRX1 | 342 | 5.77 | 37124.94 |
| Niacin | *Sorghum* | 2.4.2.19 | C5X7Q7 | 375 | 8.65 | 40261.56 |
| Niacin | *Saccharomyces* | 2.4.2.19 | P43619 | 295 | 5.55 | 32364.96 |
| Niacin | *Human* | 2.4.2.19 | Q15274 | 297 | 5.81 | 30845.63 |
| Pantothenate | *Glossina* | 2.6.1.42 | TMP003476 | 450 | 8.48 | 50779.37 |
| Pantothenate | *Glossina* | 2.6.1.42 | TMPEC2.6.1.42-641671 | 116 | 5.19 | 12877.94 |
| Pantothenate | *Drosophilla* | 2.6.1.42 | Q9VYD5 | 443 | 6.61 | 49536.87 |
| Pantothenate | *Tribolium* | 2.6.1.42 | D6WAG1 | 417 | 7.66 | 47211.36 |
| Pantothenate | *Anopheles* | 2.6.1.42 | Q7QEI9 | 443 | 6.82 | 50155.74 |
| Pantothenate | *Sodalis* | 2.6.1.42 | Q2NQA5 | 309 | 5.86 | 33649.44 |
| Pantothenate | *Escherichia* | 2.6.1.42 | P0AB80 | 309 | 5.54 | 33962.46 |
| Pantothenate | *Bacillus* | 2.6.1.42 | O31461 | 356 | 5.14 | 39676.23 |
| Pantothenate | *Bacillus* | 2.6.1.42 | P39576 | 363 | 5.13 | 40160.9 |
| Pantothenate | *Arabidopsis* | 2.6.1.42 | Q93Y32 | 384 | 5.78 | 39874.43 |
| Pantothenate | *Arabidopsis* | 2.6.1.42 | Q9M439 | 388 | 6.94 | 40023.69 |

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---|---|---|---|---|---|---|
| Pantothenate | *Arabidopsis* | 2.6.1.42 | Q9LPM8 | 367 | 5.69 | 39998.6 |
| Pantothenate | *Arabidopsis* | 2.6.1.42 | Q9LPM9 | 356 | 6.2 | 38861.43 |
| Pantothenate | *Arabidopsis* | 2.6.1.88 | Q9LE06 | 354 | 5.9 | 39018.71 |
| Pantothenate | *Arabidopsis* | 2.6.1.42 | Q9M401 | 413 | 5.77 | 38529.21 |
| Pantothenate | *Arabidopsis* | 2.6.1.42 | Q8L493 | 373 | 6.13 | 34463.2 |
| Pantothenate | *Arabidopsis* | 2.6.1.42 | Q9FYA6 | 415 | 5.24 | 38321.79 |
| Pantothenate | *Sorghum* | 2.6.1.42 | C5XZZ4 | 402 | 6.37 | 42901.09 |
| Pantothenate | *Sorghum* | 2.6.1.42 | C5YDX1 | 435 | 8.86 | 47423.44 |
| Pantothenate | *Sorghum* | 2.6.1.42 | C5YVA1 | 330 | 5.56 | 35292.54 |
| Pantothenate | *Saccharomyces* | 2.6.1.42 | P38891 | 393 | 8.48 | 41698.74 |
| Pantothenate | *Saccharomyces* | 2.6.1.42 | P47176 | 376 | 6.91 | 41624.72 |
| Pantothenate | *Human* | 2.6.1.42 | P54687 | 386 | 5.17 | 42966.15 |
| Pantothenate | *Human* | 2.6.1.42 | O15382 | 392 | 8.21 | 41264.03 |
| Pantothenate | *Wigglesworthia* | 2.1.2.11 | H6Q5V2 | 264 | 9.16 | 29256.36 |
| Pantothenate | *Sodalis* | 2.1.2.11 | Q2NVR2 | 262 | 5.66 | 27817.95 |
| Pantothenate | *Escherichia* | 2.1.2.11 | P31057 | 264 | 5.15 | 28374.44 |
| Pantothenate | *Bacillus* | 2.1.2.11 | P52996 | 277 | 5.4 | 29758.4 |
| Pantothenate | *Arabidopsis* | 2.1.2.11 | O82357 | 347 | 8.6 | 36693.33 |
| Pantothenate | *Arabidopsis* | 2.1.2.11 | Q9M315 | 354 | 6.86 | 37363.86 |
| Pantothenate | *Sorghum* | 2.1.2.11 | C5XKA7 | 404 | 8.24 | 41553.38 |
| Pantothenate | *Sorghum* | 2.1.2.11 | C5XKA8 | 370 | 6.36 | 38127.59 |
| Pantothenate | *Sorghum* | 2.1.2.11 | C5XKA9 | 353 | 8.14 | 36885.3 |
| Pantothenate | *Saccharomyces* | 2.1.2.11 | P38122 | 312 | 8.35 | 34464.91 |
| Pantothenate | *Wigglesworthia* | 1.1.1.169 | H6Q5G7 | 294 | 9.7 | 34632.38 |
| Pantothenate | *Sodalis* | 1.1.1.169 | Q2NV89 | 303 | 8.74 | 33752.66 |
| Pantothenate | *Escherichia* | 1.1.1.169 | P0A9J4 | 303 | 5.62 | 33870.76 |
| Pantothenate | *Bacillus* | 1.1.1.- | O31717 | 303 | 6.14 | 33572.42 |
| Pantothenate | *Bacillus* | 1.1.1.- | O34661 | 298 | 5.88 | 33287.04 |
| Pantothenate | *Saccharomyces* | 1.1.1.169 | P38787 | 379 | 6.21 | 42821.28 |
| Pantothenate | *Glossina* | 1.3.1.2 | TMP002436 | 1035 | 7.52 | 112669.9 |
| Pantothenate | *Drosophilla* | 1.3.1.2; 1.3.1.1; 1.3.3.1 | Q9W374 | 1031 | 6.42 | 111225.6 |
| Pantothenate | *Tribolium* | 1.3.1.2 | D6WGA9 | 1016 | 8.13 | 109951.8 |
| Pantothenate | *Anopheles* | 1.3.1.2 | Q7QHY0 | 1039 | 6.72 | 112361.1 |
| Pantothenate | *Arabidopsis* | 1.3.1.2 | Q9LVI9 | 426 | 6.37 | 46846.61 |
| Pantothenate | *Sorghum* | 1.3.1.2 | C5XZ92 | 420 | 5.92 | 45922.58 |
| Pantothenate | *Human* | 1.3.1.2 | Q12882 | 1025 | 6.84 | 111102 |
| Pantothenate | *Glossina* | 3.5.2.2 | TMP006803 | 594 | 6.13 | 65129.07 |
| Pantothenate | *Drosophilla* | 3.5.2.2 | Q8IPQ2 | 594 | 6.23 | 65101.28 |
| Pantothenate | *Tribolium* | 3.5.2.2 | 662209/XP_973416.2 | 587 | 5.94 | 64851.84 |
| Pantothenate | *Anopheles* | 3.5.2.2 | Q7QBK9 | 333 | 6.3 | 65415.12 |
| Pantothenate | *Escherichia* | 3.5.2.- | Q46806 | 461 | 6.18 | 51025.15 |
| Pantothenate | *Arabidopsis* | 3.5.2.2 | Q9FMP3 | 531 | 5.58 | 57991.22 |
| Pantothenate | *Sorghum* | 3.5.2.2 | C5XMK0 | 536 | 6.02 | 57550.93 |
| Pantothenate | *Human* | 3.5.2.2 | Q14117 | 519 | 6.81 | 56629.78 |
| Pantothenate | *Glossina* | 3.5.1.6 | TMP008479 | 386 | 6.84 | 43771.66 |
| Pantothenate | *Drosophilla* | 3.5.1.6 | Q9VI04 | 386 | 6.4 | 43799.76 |
| Pantothenate | *Tribolium* | 3.5.1.6 | D2A4C0 | 383 | 6.28 | 43450.4 |
| Pantothenate | *Anopheles* | 3.5.1.6 | Q7Q0P4 | 386 | 6.03 | 43520.4 |
| Pantothenate | *Arabidopsis* | 3.5.1.6 | Q8H183 | 408 | 5.92 | 45552.61 |
| Pantothenate | *Sorghum* | 3.5.1.6 | C5X8L4 | 413 | 5.86 | 45686.84 |
| Pantothenate | *Human* | 3.5.1.6 | Q9UBR1 | 384 | 6.09 | 43166.07 |
| Pantothenate | *Sodalis* | 4.1.1.11 | Q2NVR4 | 126 | 6.02 | 14089.01 |
| Pantothenate | *Escherichia* | 4.1.1.11 | P0A790 | 126 | 5.75 | 13833.73 |
| Pantothenate | *Bacillus* | 4.1.1.11 | P52999 | 127 | 5.65 | 13899.98 |
| Pantothenate | *Wigglesworthia* | 6.3.2.1 | H6Q5V3 | 290 | 9.85 | 33654.15 |
| Pantothenate | *Sodalis* | 6.3.2.1 | Q2NVR3 | 284 | 5.79 | 31705.48 |
| Pantothenate | *Escherichia* | 6.3.2.1 | P31663 | 283 | 5.92 | 31597.67 |
| Pantothenate | *Bacillus* | 6.3.2.1 | P52998 | 286 | 4.82 | 31958.34 |
| Pantothenate | *Arabidopsis* | 6.3.2.1 | Q9FKB3 | 310 | 6.05 | 34137.07 |
| Pantothenate | *Sorghum* | 6.3.2.1 | C5WS23 | 315 | 5.45 | 34062.76 |
| Pantothenate | *Saccharomyces* | 6.3.2.1 | P40459 | 309 | 5.27 | 35032.24 |
| Pantothenate | *Glossina* | 2.7.1.33 | TMP008185 | 391 | 6.25 | 44714.39 |
| Pantothenate | *Glossina* | 2.7.1.33 | TMP002949 | 495 | 8.94 | 55586.42 |
| Pantothenate | *Drosophilla* | 2.7.1.33 | D8FT20 | 342 | 6.02 | 37672.79 |
| Pantothenate | *Drosophilla* | 2.7.1.33 | Q9VMU2 | 361 | 5.36 | 40442.2 |
| Pantothenate | *Tribolium* | 2.7.1.33 | D6WCF9 | 367 | 7.51 | 41020.01 |
| Pantothenate | *Tribolium* | 2.7.1.33 | D6WMP7 | 355 | 5.18 | 40072.85 |
| Pantothenate | *Anopheles* | 2.7.1.33 | Q7Q0Y3 | 372 | 5.82 | 42210.46 |
| Pantothenate | *Wigglesworthia* | 2.7.1.33 | H6Q4M3 | 272 | 9.8 | 32222.29 |
| Pantothenate | *Sodalis* | 2.7.1.33 | Q2NWS4 | 316 | 7.92 | 35856.92 |
| Pantothenate | *Escherichia* | 2.7.1.33 | P0A6I3 | 316 | 6.32 | 36359.78 |
| Pantothenate | *Bacillus* | 2.7.1.33 | P37564 | 258 | 5.89 | 28576.15 |
| Pantothenate | *Bacillus* | 2.7.1.33 | P54556 | 319 | 5.92 | 36639.75 |

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---|---|---|---|---|---|---|
| Pantothenate | *Arabidopsis* | 2.7.1.33 | Q8L5Y9 | 901 | 5.6 | 99639.65 |
| Pantothenate | *Sorghum* | 2.7.1.33 | C5X682 | 902 | 5.67 | 99469.79 |
| Pantothenate | *Saccharomyces* | 2.7.1.33 | Q04430 | 367 | 5.93 | 40903.17 |
| Pantothenate | *Human* | 2.7.1.33 | Q8TE04 | 598 | 7.51 | 64339.34 |
| Pantothenate | *Human* | 2.7.1.33 | Q9NVE7 | 773 | 5.87 | 85990.95 |
| Pantothenate | *Human* | 2.7.1.33 | Q9H999 | 370 | 6.13 | 41094.07 |
| Pantothenate | *Human* | 2.7.1.33 | Q9BZ23 | 570 | 9.05 | 57488.55 |
| Pantothenate | *Glossina* | 6.3.2.5 | TMP004399 | 306 | 6.5 | 35272.39 |
| Pantothenate | *Drosophilla* | 6.3.2.5 | Q7KN99 | 313 | 5.85 | 35751.85 |
| Pantothenate | *Tribolium* | 6.3.2.5 | D6X537 | 311 | 6.39 | 35308.63 |
| Pantothenate | *Anopheles* | 6.3.2.5 | Q7QAC3 | 316 | 6.24 | 36103.29 |
| Pantothenate | *Wigglesworthia* | 6.3.2.5/4.1.1.36 | H6Q4R3 | 407 | 9.65 | 45181.7 |
| Pantothenate | *Sodalis* | 6.3.2.5/4.1.1.36 | Q2NQU1 | 404 | 6 | 43157.56 |
| Pantothenate | *Escherichia* | 4.1.1.36; 6.3.2.5 | P0ABQ0 | 406 | 7.04 | 43306.95 |
| Pantothenate | *Bacillus* | 4.1.1.36; 6.3.2.5 | O35033 | 406 | 6.04 | 43977.75 |
| Pantothenate | *Arabidopsis* | 6.3.2.5 | Q8GXR5 | 317 | 7.54 | 35368.8 |
| Pantothenate | *Arabidopsis* | 6.3.2.5 | Q9LZM3 | 309 | 6.34 | 34823.91 |
| Pantothenate | *Sorghum* | 6.3.2.5 | C5XVP2 | 332 | 8.66 | 36820.29 |
| Pantothenate | *Sorghum* | 6.3.2.5 | C5YHW1 | 321 | 7.05 | 35629.07 |
| Pantothenate | *Saccharomyces* | 6.3.2.5 | P40506 | 365 | 8.26 | 41866.95 |
| Pantothenate | *Human* | 6.3.2.5 | Q9HAB8 | 311 | 6.26 | 34005.13 |
| Pantothenate | *Glossina* | 4.1.1.36 | TMP011430 | 189 | 6.2 | 21115.92 |
| Pantothenate | *Glossina* | 4.1.1.36 | TMPEC4.1.1.36-651746 | 55 | 5.16 | 6675.68 |
| Pantothenate | *Drosophilla* | 4.1.1.36 | Q8MKK3 | 191 | 6.21 | 21542.24 |
| Pantothenate | *Tribolium* | 4.1.1.36 | D2CFX2 | 188 | 6.04 | 20755.32 |
| Pantothenate | *Anopheles* | 4.1.1.36 | Q7PZN2 | 191 | 5.77 | 21124.66 |
| Pantothenate | *Wigglesworthia* | 4.1.1.36 | H6Q4R3 | 407 | 9.65 | 45181.7 |
| Pantothenate | *Sodalis* | 4.1.1.36 | Q2NQU1 | 404 | 6 | 43157.56 |
| Pantothenate | *Escherichia* | 4.1.1.36 | P0ABQ0 | 406 | 7.04 | 43306.95 |
| Pantothenate | *Bacillus* | 4.1.1.36 | O35033 | 406 | 6.04 | 43977.75 |
| Pantothenate | *Arabidopsis* | 4.1.1.36 | P94063 | 201 | 6.82 | 22414.94 |
| Pantothenate | *Arabidopsis* | 4.1.1.36 | Q9SWE5 | 209 | 6.21 | 23354.87 |
| Pantothenate | *Sorghum* | 4.1.1.36 | C5XLV1 | 200 | 6.51 | 21777.04 |
| Pantothenate | *Sorghum* | 4.1.1.36 | C5Z606 | 193 | 5.46 | 21705.29 |
| Pantothenate | *Sorghum* | 4.1.1.36 | C5Z607 | 220 | 5.94 | 24224.71 |
| Pantothenate | *Saccharomyces* | 4.1.1.36 | P36076 | 571 | 4.91 | 65238.26 |
| Pantothenate | *Human* | 4.1.1.36 | Q96CD2 | 204 | 5.72 | 22395.18 |
| Pantothenate | *Glossina* | 2.7.7.3/2.7.1.24 | TMP002599 | 553 | 7.62 | 62847.5 |
| Pantothenate | *Glossina* | 2.7.7.3/2.7.1.24 | TMP009026 | 240 | 9.13 | 27711.45 |
| Pantothenate | *Drosophilla* | 2.7.1.24; 2.7.7.-; 2.7.7.3 | Q9VRP4 | 518 | 6.72 | 57641.15 |
| Pantothenate | *Tribolium* | 2.7.7.3/2.7.1.24 | 664060/XP_975172.1 | 513 | 6.17 | 57291.94 |
| Pantothenate | *Anopheles* | ^2.7.7.3/2.7.1.24 | Q7Q774 | 521 | 6.18 | 58338.73 |
| Pantothenate | *Wigglesworthia* | 2.7.7.3 | H6Q528 | 161 | 9.76 | 18853.42 |
| Pantothenate | *Sodalis* | 2.7.7.3 | Q2NQU5 | 160 | 7.02 | 17661.6 |
| Pantothenate | *Wolbachia* | 2.7.7.3 | Q73HM7 | 168 | 5.46 | 18828.66 |
| Pantothenate | *Escherichia* | 2.7.7.3 | P0A6I6 | 159 | 6.49 | 17836.63 |
| Pantothenate | *Bacillus* | 2.7.7.3 | O34797 | 161 | 5.75 | 18178.8 |
| Pantothenate | *Arabidopsis* | 2.7.7.3 | Q9ZPV8 | 176 | 6.19 | 19167.95 |
| Pantothenate | *Sorghum* | 2.7.7.3 | C5XB13 | 188 | 5.13 | 20182.95 |
| Pantothenate | *Sorghum* | 2.7.7.3 | C5XTW9 | 225 | 5.47 | 23986.26 |
| Pantothenate | *Saccharomyces* | 2.7.7.3 | P53332 | 305 | 5.89 | 34306.52 |
| Pantothenate | *Human* | 2.7.7.3; 2.7.1.24 | Q13057 | 564 | 6.51 | 62328.84 |
| Pantothenate | *Glossina* | 2.7.7.3/2.7.1.24 | TMP002599 | 553 | 7.62 | 62847.5 |
| Pantothenate | *Glossina* | 2.7.7.3/2.7.1.24 | TMP009026 | 240 | 9.13 | 27711.45 |
| Pantothenate | *Drosophilla* | 2.7.1.24; 2.7.7.-; 2.7.7.3 | Q9VRP4 | 518 | 6.72 | 57641.15 |
| Pantothenate | *Tribolium* | 2.7.7.3/2.7.1.24 | 664060/XP_975172.1 | 513 | 6.17 | 57291.94 |
| Pantothenate | *Anopheles* | 2.7.7.3/2.7.1.24 | Q7Q774 | 521 | 6.18 | 58338.73 |
| Pantothenate | *Wigglesworthia* | 2.7.1.24 | H6Q5C1 | 195 | 9.97 | 23135.52 |
| Pantothenate | *Sodalis* | 2.7.1.24 | Q2NVT9 | 208 | 9.41 | 22876.07 |
| Pantothenate | *Wolbachia* | 2.7.1.24 | Q73IH9 | 195 | 5.28 | 22297.74 |
| Pantothenate | *Escherichia* | 2.7.1.24 | P0A6I9 | 206 | 5.77 | 22621.71 |
| Pantothenate | *Bacillus* | 2.7.1.24 | O34932 | 197 | 5.05 | 22013.44 |
| Pantothenate | *Arabidopsis* | 2.7.1.24 | Q9ZQH0 | 232 | 9.54 | 25747.09 |
| Pantothenate | *Sorghum* | 2.7.1.24 | C5XKI1 | 232 | 9.51 | 26170.38 |
| Pantothenate | *Sorghum* | 2.7.1.24 | C5YLK2 | 230 | 9.17 | 25634.9 |
| Pantothenate | *Saccharomyces* | 2.7.1.24 | Q03941 | 241 | 8.76 | 27339.76 |
| Pantothenate | *Human* | 2.7.1.24 | Q13057 | 564 | 6.51 | 62328.84 |
| Folate | *Glossina* | 3.5.4.16 | TMP011539 | 384 | 6.49 | 42590.79 |
| Folate | *Drosophilla* | 3.5.4.16 | P48596 | 324 | 6.16 | 35541.46 |
| Folate | *Tribolium* | 3.5.4.16 | D2CFW6 | 256 | 7.11 | 28824.27 |
| Folate | *Anopheles* | 3.5.4.16 | A7UU97 | 285 | 6.86 | 31486.81 |
| Folate | *Wigglesworthia* | 3.5.4.16 | H6Q500 | 221 | 9.95 | 25581.22 |
| Folate | *Sodalis* | 3.5.4.16 | Q2NUE3 | 221 | 7.06 | 24829.71 |

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---------|----------|-----------|-------------|--------|-------------------|------------------|
| Folate | *Escherichia* | 3.5.4.16 | P0A6T5 | 222 | 6.93 | 24699.43 |
| Folate | *Bacillus* | 3.5.4.16 | P19465 | 190 | 6.31 | 21219.41 |
| Folate | *Arabidopsis* | 3.5.4.16 | F4JED5 | 466 | 6.96 | 51483.25 |
| Folate | *Saccharomyces* | 3.5.4.16 | P51601 | 243 | 6.72 | 27769.05 |
| Folate | *Human* | 3.5.4.16 | P30793 | 250 | 8.73 | 27903.07 |
| Folate | *Glossina* | 3.1.3.1 | TMP006886 | 539 | 6.44 | 59616.45 |
| Folate | *Glossina* | 3.1.3.1 | TMP002067 | 547 | 6.26 | 60537.32 |
| Folate | *Glossina* | 3.1.3.1 | TMP012132 | 526 | 5.08 | 58599.75 |
| Folate | *Glossina* | 3.1.3.1 | TMP009009 | 541 | 5.74 | 60217.31 |
| Folate | *Glossina* | 3.1.3.1 | TMP006977 | 589 | 6.19 | 65269.87 |
| Folate | *Glossina* | 3.1.3.1 | TMP009468 | 611 | 6.58 | 67402.05 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9VRM8 | 524 | 5.17 | 58140.67 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9VDG4 | 522 | 5.04 | 57295.03 |
| Folate | *Drosophilla* | 3.1.3.1 | Q24238 | 596 | 5.71 | 60657.64 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9VIW9 | 596 | 5.25 | 65470.27 |
| Folate | *Drosophilla* | 3.1.3.1 | Q7K3X8 | 515 | 5.33 | 56545.88 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9W273 | 533 | 4.93 | 57407.37 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9W274 | 538 | 5.76 | 58227.48 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9W275 | 543 | 4.97 | 58258.28 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9VRM9 | 517 | 5.53 | 56641.84 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9VH28 | 450 | 5.15 | 50530.71 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9VP35 | 523 | 6.21 | 57646.58 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9VXS8 | 483 | 6.12 | 52593.5 |
| Folate | *Drosophilla* | 3.1.3.1 | Q9VHD0 | 546 | 5.95 | 59504.61 |
| Folate | *Tribolium* | 3.1.3.1 | 657368/XP_968925.2 | 520 | 6.29 | 57575.83 |
| Folate | *Tribolium* | 3.1.3.1 | 660001/XP_971358.2 | 508 | 5.41 | 55499.41 |
| Folate | *Tribolium* | 3.1.3.1 | D6WY51 | 529 | 5.43 | 58438.26 |
| Folate | *Tribolium* | 3.1.3.1 | D6WY52 | 503 | 5.32 | 55131.34 |
| Folate | *Tribolium* | 3.1.3.1 | 661869/XP_973094.1 | 574 | 6.68 | 62608.74 |
| Folate | *Tribolium* | 3.1.3.1 | D2A3C2 | 545 | 5.24 | 60610.26 |
| Folate | *Anopheles* | 3.1.3.1 | Q7PY02 | 548 | 5.82 | 60708.6 |
| Folate | *Anopheles* | 3.1.3.1 | Q7QJ58 | 571 | 6.06 | 62203.09 |
| Folate | *Anopheles* | 3.1.3.1 | Q7Q8R8 | 574 | 6.33 | 63339.72 |
| Folate | *Anopheles* | 3.1.3.1 | Q7QHQ7 | 531 | 5.47 | 59191.22 |
| Folate | *Anopheles* | 3.1.3.1 | A7US49 | 571 | 5.26 | 62845.26 |
| Folate | *Escherichia* | 3.1.3.1 | P00634 | 471 | 5.54 | 47199.79 |
| Folate | *Bacillus* | 3.1.3.1 | P42251 | 583 | 8.63 | 59704.72 |
| Folate | *Bacillus* | 3.1.3.1 | P19405 | 462 | 5.62 | 47235.85 |
| Folate | *Bacillus* | 3.1.3.1 | P19406 | 461 | 9.39 | 45939 |
| Folate | *Arabidopsis* | 3.1.3.1 | F4K1J1 | 453 | 8.99 | 51131.87 |
| Folate | *Sorghum* | 3.1.3.1 | C5XWF9 | 449 | 8.89 | 50787.35 |
| Folate | *Saccharomyces* | 3.1.3.1; 3.1.7.6 | P11491 | 566 | 5.31 | 63004.11 |
| Folate | *Human* | 3.1.3.1 | P09923 | 528 | 5.39 | 52444.81 |
| Folate | *Human* | 3.1.3.1 | P05186 | 524 | 6.19 | 53390.01 |
| Folate | *Human* | 3.1.3.1 | P05187 | 535 | 5.73 | 52745.31 |
| Folate | *Human* | 3.1.3.1 | P10696 | 532 | 5.78 | 52573.1 |
| Folate | *Wigglesworthia* | 4.1.2.25 | H6Q5S8 | 120 | 9.33 | 13843.57 |
| Folate | *Sodalis* | 4.1.2.25 | Q2NWE4 | 119 | 5.39 | 13502.74 |
| Folate | *Escherichia* | 4.1.2.25 | P0AC16 | 122 | 4.68 | 13619.53 |
| Folate | *Bacillus* | 4.1.2.25 | P28823 | 120 | 5.37 | 13516.5 |
| Folate | *Arabidopsis* | 4.1.2.25 | Q9SF23 | 146 | 6.31 | 16280.65 |
| Folate | *Arabidopsis* | 4.1.2.25 | F4IYU3 | 156 | 6.31 | 17630.56 |
| Folate | *Arabidopsis* | 4.1.2.25 | Q9FM54 | 131 | 5.85 | 14685.19 |
| Folate | *Sorghum* | 4.1.2.25 | C5YNA8 | 133 | 5.82 | 14378.54 |
| Folate | *Saccharomyces* | 4.1.2.25; 2.7.6.3; 2.5.1.15 | P53848 | 824 | 6.01 | 93119.94 |
| Folate | *Glossina* | 2.7.6.3 | TMPEC2.7.6.3-641167 | 360 | 4.94 | 40108.18 |
| Folate | *Glossina* | 2.7.6.3 | TMPEC2.7.6.3-642052 | 92 | 8.64 | 10547.27 |
| Folate | *Wigglesworthia* | 2.7.6.3 | H6Q551 | 161 | 9.75 | 18797.04 |
| Folate | *Sodalis* | 2.7.6.3 | Q2NVR1 | 168 | 5.48 | 19020.66 |
| Folate | *Wolbachia* | 2.7.6.3/2.5.1.15 | Q73GQ3 | 198 | 5.54 | 22240.33 |
| Folate | *Escherichia* | 2.7.6.3 | P26281 | 159 | 5.34 | 17947.6 |
| Folate | *Bacillus* | 2.7.6.3 | P29252 | 167 | 4.73 | 19058.46 |
| Folate | *Arabidopsis* | 2.5.1.15; 2.7.6.3 | Q1ENB6 | 484 | 5.79 | 54110.34 |
| Folate | *Arabidopsis* | 2.5.1.15; 2.7.6.3 | F4JPH1 | 561 | 6.42 | 61426.53 |
| Folate | *Sorghum* | 2.7.6.3/2.5.1.15 | C5X2E7 | 505 | 9.04 | 55233.78 |
| Folate | *Sorghum* | 2.7.6.3/2.5.1.15 | C5XIR9 | 615 | 8.87 | 66086.63 |
| Folate | *Saccharomyces* | 4.1.2.25; 2.7.6.3; 2.5.1.15 | P53848 | 824 | 6.01 | 93119.94 |
| Folate | *Wigglesworthia* | 2.6.1.85 | H6Q5R8 | 459 | 9.43 | 52696.5 |
| Folate | *Sodalis* | 2.6.1.85 | Q2NTC1 | 456 | 5.58 | 49912.04 |
| Folate | *Sodalis* | 2.6.1.85 | Q2NQJ8 | 193 | 6.75 | 21335.53 |
| Folate | *Escherichia* | 2.6.1.85 | P05041 | 453 | 5.08 | 50969.55 |
| Folate | *Escherichia* | 2.6.1.85 | P00903 | 187 | 6.11 | 20771.86 |
| Folate | *Bacillus* | 2.6.1.85 | P28820 | 470 | 5.37 | 53251.2 |

105

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---|---|---|---|---|---|---|
| Folate | *Bacillus* | 2.6.1.85; 4.1.3.27 | P28819 | 194 | 4.96 | 21684.9 |
| Folate | *Arabidopsis* | 2.6.1.85 | Q8LPN3 | 919 | 6.18 | 102919.7 |
| Folate | *Sorghum* | 2.6.1.85 | C5Z8W2 | 858 | 6.54 | 95511.84 |
| Folate | *Saccharomyces* | 2.6.1.85 | P37254 | 787 | 5.52 | 88544.14 |
| Folate | *Wigglesworthia* | 4.1.3.38 | H6Q5Q3 | 267 | 10.09 | 31323.26 |
| Folate | *Sodalis* | 4.1.3.38 | Q2NU37 | 265 | 7.1 | 29179.58 |
| Folate | *Escherichia* | 4.1.3.38 | P28305 | 269 | 6.08 | 29715.06 |
| Folate | *Bacillus* | 4.1.3.38 | P28821 | 293 | 6.14 | 33515.46 |
| Folate | *Wigglesworthia* | 2.5.1.15 | H6Q587 | 280 | 8.83 | 31617.16 |
| Folate | *Sodalis* | 2.5.1.15 | Q2NW28 | 281 | 5.59 | 29929.56 |
| Folate | *Wolbachia* | 2.5.1.15/2.7.6.3 | Q73GQ3 | 198 | 5.54 | 22240.33 |
| Folate | *Escherichia* | 2.5.1.15 | P0AC13 | 282 | 5.67 | 30615.11 |
| Folate | *Bacillus* | 2.5.1.15 | P28822 | 285 | 5.46 | 31001.64 |
| Folate | *Arabidopsis* | 2.5.1.15; 2.7.6.3 | Q1ENB6 | 484 | 5.79 | 54110.34 |
| Folate | *Arabidopsis* | 2.5.1.15; 2.7.6.3 | F4JPH1 | 561 | 6.42 | 61426.53 |
| Folate | *Sorghum* | 2.5.1.15/2.7.6.3 | C5X2E7 | 505 | 9.04 | 55233.78 |
| Folate | *Sorghum* | 2.5.1.15/2.7.6.3 | C5XIR9 | 615 | 8.87 | 66086.63 |
| Folate | *Saccharomyces* | 4.1.2.25; 2.7.6.3; 2.5.1.15 | P53848 | 824 | 6.01 | 93119.94 |
| Folate | *Glossina* | 6.3.2.17 | TMP007575 | 524 | 7.57 | 59985.72 |
| Folate | *Glossina* | 6.3.2.17 | TMP007308 | 310 | 6.85 | 35371.33 |
| Folate | *Glossina* | 6.3.2.17 | TMP002232 | 333 | 7.56 | 37937.5 |
| Folate | *Glossina* | 6.3.2.17 | TMPEC6.3.2.17-641167 | 303 | 5.58 | 33397.6 |
| Folate | *Drosophilla* | 6.3.2.17 | Q9VYL1 | 572 | 7.65 | 64317.16 |
| Folate | *Drosophilla* | 6.3.2.17 | Q9VQW0 | 758 | 9.43 | 84355.99 |
| Folate | *Tribolium* | 6.3.2.17 | D6W9Q8 | 464 | 8.3 | 52504.32 |
| Folate | *Anopheles* | 6.3.2.17 | Q5TSB4 | 530 | 8.11 | 60369.45 |
| Folate | *Wigglesworthia* | 6.3.2.12/6.3.2.17 | H6Q506 | 418 | 9.74 | 47621.09 |
| Folate | *Sodalis* | 6.3.2.12/6.3.2.17 | Q2NSI4 | 421 | 5.85 | 45252.82 |
| Folate | *Wolbachia* | 6.3.2.17 | Q73GA7 | 429 | 6.33 | 47370.65 |
| Folate | *Escherichia* | 6.3.2.17; 6.3.2.12 | P08192 | 422 | 5.5 | 45405.7 |
| Folate | *Bacillus* | 6.3.2.17 | Q05865 | 430 | 5.84 | 48164.89 |
| Folate | *Arabidopsis* | 6.3.2.17 | F4J2K2 | 625 | 8.69 | 68920.64 |
| Folate | *Arabidopsis* | 6.3.2.17 | Q8W035 | 492 | 7.55 | 55167.32 |
| Folate | *Arabidopsis* | 6.3.2.17 | F4K2A1 | 571 | 6.08 | 63345.39 |
| Folate | *Arabidopsis* | 6.3.2.17 | F4JYE9 | 530 | 6.72 | 56904.06 |
| Folate | *Sorghum* | 6.3.2.17 | C5WWE5 | 588 | 7.05 | 64578.87 |
| Folate | *Saccharomyces* | 6.3.2.17 | P36001 | 430 | 6.75 | 48143.42 |
| Folate | *Saccharomyces* | 6.3.2.17 | Q12676 | 427 | 6.36 | 47851.15 |
| Folate | *Saccharomyces* | 6.3.2.17 | Q08645 | 548 | 8.92 | 62151.44 |
| Folate | *Human* | 6.3.2.17 | Q05932 | 587 | 6.85 | 60166.97 |
| Folate | *Glossina* | 1.5.1.3 | TMP010615 | 343 | 7.81 | 38347.4 |
| Folate | *Glossina* | 1.5.1.3 | TMPEC1.5.1.3-639395 | 187 | 8.97 | 21469.04 |
| Folate | *Glossina* | 1.5.1.3 | TMPEC1.5.1.3-641167 | 159 | 9.15 | 18265.45 |
| Folate | *Drosophilla* | 1.5.1.3 | P17719 | 182 | 6.18 | 20775.01 |
| Folate | *Anopheles* | 1.5.1.3 | Q7Q0L5 | 187 | 6.22 | 21376.31 |
| Folate | *Wigglesworthia* | 1.5.1.3 | H6Q4A2 | 166 | 9.76 | 19821.37 |
| Folate | *Sodalis* | 1.5.1.3 | Q2NVX9 | 161 | 4.99 | 18019.44 |
| Folate | *Wolbachia* | 1.5.1.3 | Q73GQ2 | 159 | 8.75 | 18237.39 |
| Folate | *Escherichia* | 1.5.1.3 | P0ABQ4 | 159 | 4.84 | 17999.38 |
| Folate | *Bacillus* | 1.5.1.3 | P11045 | 168 | 5.35 | 19175.68 |
| Folate | *Arabidopsis* | 1.5.1.3; 2.1.1.45 | Q05762 | 519 | 5.55 | 58143.4 |
| Folate | *Arabidopsis* | 1.5.1.3; 2.1.1.45 | Q05763 | 565 | 8.13 | 63208.68 |
| Folate | *Sorghum* | 1.5.1.3/2.5.1.15 | C5Y2E9 | 521 | 6.05 | 58819.23 |
| Folate | *Saccharomyces* | 1.5.1.3 | P07807 | 211 | 7.67 | 24261 |
| Folate | *Human* | 1.5.1.3 | P00374 | 187 | 7.01 | 21321.54 |
| Folate | *Human* | 1.5.1.3 | Q86XF0 | 187 | 7.75 | 21620.05 |
| Biotin | *Wigglesworthia* | BioC | H6Q5U6 | 262 | 9.35 | 30511.07 |
| Biotin | *Sodalis* | BioC | Q2NUJ5 | 259 | 9.79 | 28764.89 |
| Biotin | *Escherichia* | BioC | P12999 | 251 | 7.93 | 28276.2 |
| Biotin | *Wigglesworthia* | FabB | H6Q5Z4 | 404 | 8.99 | 43437 |
| Biotin | *Sodalis* | FabB | Q2NSH7 | 403 | 5.42 | 42476.25 |
| Biotin | *Escherichia* | FabB | P0A953 | 406 | 5.34 | 42613.32 |
| Biotin | *Glossina* | 2.3.1.41 | TMP003897 | 423 | 6.54 | 45135.38 |
| Biotin | *Glossina* | 2.3.1.41 | TMP007107 | 449 | 8.65 | 48384.53 |
| Biotin | *Drosophilla* | 2.3.1.41 | Q9VNF5 | 438 | 8.04 | 46378.86 |
| Biotin | *Tribolium* | 2.3.1.41 | D6X1D1 | 421 | 7.57 | 43887.2 |
| Biotin | *Anopheles* | 2.3.1.41 | Q7QCU6 | 459 | 7.61 | 48966.85 |
| Biotin | *Sodalis* | 2.3.1.179 | Q2NU38 | 413 | 5.82 | 43032.69 |
| Biotin | *Wolbachia* | 2.3.1.179 | Q73FX9 | 423 | 6.28 | 45200.41 |
| Biotin | *Escherichia* | 2.3.1.179 | P0AAI5 | 413 | 5.71 | 42914.57 |
| Biotin | *Bacillus* | 2.3.1.179 | O34340 | 413 | 4.95 | 44004.73 |
| Biotin | *Arabidopsis* | 2.3.1.41 | Q9C9P4 | 541 | 6.09 | 46759.48 |
| Biotin | *Arabidopsis* | 2.3.1.41 | Q8L3X9 | 461 | 6.08 | 46134.41 |

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---------|----------|-----------|-------------|--------|-------------------|------------------|
| Biotin | *Sorghum* | 2.3.1.41 | C5XE14 | 496 | 7.84 | 53031.88 |
| Biotin | *Sorghum* | 2.3.1.41 | C5X456 | 471 | 8.55 | 49204.93 |
| Biotin | *Sorghum* | 2.3.1.41 | C5XXH6 | 464 | 6.65 | 48360.82 |
| Biotin | *Sorghum* | 2.3.1.41 | C5Z5Z9 | 461 | 7.07 | 48710.56 |
| Biotin | *Saccharomyces* | 2.3.1.41 | P39525 | 442 | 8.44 | 47555.21 |
| Biotin | *Human* | 2.3.1.41 | Q9NWU1 | 459 | 6.53 | 45703.12 |
| Biotin | *Wigglesworthia* | 1.1.1.100 | H6Q5Q1 | 244 | 9.71 | 26717.27 |
| Biotin | *Sodalis* | 1.1.1.100 | Q2NU40 | 244 | 6.18 | 25618.44 |
| Biotin | *Wolbachia* | 1.1.1.100 | Q73HB7 | 244 | 8.22 | 26353.63 |
| Biotin | *Escherichia* | 1.1.1.100 | P0AEK2 | 244 | 6.76 | 25560.29 |
| Biotin | *Bacillus* | 1.1.1.100 | P51831 | 246 | 7.76 | 26282.17 |
| Biotin | *Bacillus* | 1.1.1.100 | O31767 | 242 | 5.36 | 25541.09 |
| Biotin | *Bacillus* | 1.1.1.100 | O34308 | 255 | 6.91 | 27974.13 |
| Biotin | *Arabidopsis* | 1.1.1.100 | P33207 | 319 | 8.38 | 27138.4 |
| Biotin | *Arabidopsis* | 1.1.1.100 | Q9SQR4 | 270 | 6.13 | 28146.14 |
| Biotin | *Arabidopsis* | 1.1.1.100 | Q9SQR2 | 272 | 6.84 | 28434.64 |
| Biotin | *Arabidopsis* | 1.1.1.100 | Q9SVQ9 | 263 | 6.33 | 27656.63 |
| Biotin | *Arabidopsis* | 1.1.1.100 | F4JWJ4 | 277 | 9.16 | 29206.57 |
| Biotin | *Sorghum* | 1.1.1.100 | C5WZI4 | 256 | 7.77 | 26282.91 |
| Biotin | *Sorghum* | 1.1.1.100 | C5WZI6 | 261 | 5.69 | 27029.72 |
| Biotin | *Sorghum* | 1.1.1.100 | C5WZI7 | 261 | 6.76 | 26957.68 |
| Biotin | *Sorghum* | 1.1.1.100 | C5WWL1 | 275 | 5.61 | 27788.32 |
| Biotin | *Sorghum* | 1.1.1.100 | C5XAC7 | 275 | 5.49 | 27469.1 |
| Biotin | *Sorghum* | 1.1.1.100 | C5X9U6 | 274 | 5.96 | 28185.02 |
| Biotin | *Sorghum* | 1.1.1.100 | C5XSJ4 | 316 | 9.37 | 32714.93 |
| Biotin | *Sorghum* | 1.1.1.100 | C5XUE9 | 294 | 6.06 | 29699.53 |
| Biotin | *Sorghum* | 1.1.1.100 | C5YE75 | 316 | 9.26 | 32515.5 |
| Biotin | *Glossina* | 4.2.1.59 | TMPEC4.2.1.59-650627 | 143 | 8.72 | 15747.57 |
| Biotin | *Wigglesworthia* | 4.2.1.59 | H6Q4T3 | 155 | 9.46 | 17534.44 |
| Biotin | *Sodalis* | 4.2.1.59 | Q2NRL8 | 151 | 7.81 | 17039.93 |
| Biotin | *Wolbachia* | 4.2.1.59 | P61455 | 143 | 8.36 | 15802.6 |
| Biotin | *Escherichia* | 4.2.1.59 | P0A6Q6 | 151 | 6.84 | 17032.95 |
| Biotin | *Bacillus* | 4.2.1.59 | P94584 | 141 | 5.95 | 15740.23 |
| Biotin | *Arabidopsis* | 4.2.1.59 | Q9SIE3 | 220 | 8.61 | 24242.1 |
| Biotin | *Arabidopsis* | 4.2.1.59 | Q9LX13 | 219 | 9.33 | 24123.27 |
| Biotin | *Sorghum* | 4.2.1.59 | C5YIY2 | 224 | 9.27 | 24323.48 |
| Biotin | *Sorghum* | 4.2.1.59 | C5YYP0 | 216 | 9.82 | 23780.04 |
| Biotin | *Wigglesworthia* | 1.3.1.9 | H6Q4V4 | 262 | 9.49 | 29368.91 |
| Biotin | *Sodalis* | 1.3.1.9 | Q2NST7 | 262 | 5.36 | 27906.76 |
| Biotin | *Wolbachia* | 1.3.1.9 | Q73IR6 | 261 | 5.7 | 28482.6 |
| Biotin | *Escherichia* | 1.3.1.9 | P0AEK4 | 262 | 5.58 | 27732.75 |
| Biotin | *Bacillus* | 1.3.1.9 | P54616 | 258 | 5.67 | 27873.74 |
| Biotin | *Arabidopsis* | 1.3.1.9 | Q9SLA8 | 390 | 5.85 | 33361.83 |
| Biotin | *Wigglesworthia* | 3.1.1.85 | H6Q4F0 | 259 | 10 | 29735.4 |
| Biotin | *Sodalis* | 3.1.1.85 | Q2NQH6 | 257 | 6.75 | 27985.2 |
| Biotin | *Escherichia* | 3.1.1.85 | P13001 | 256 | 6.5 | 28505.01 |
| Biotin | *Wigglesworthia* | 2.3.1.47 | H6Q5U7 | 404 | 9.91 | 46304.13 |
| Biotin | *Sodalis* | 2.3.1.47 | Q2NUJ6 | 388 | 7.82 | 41762.51 |
| Biotin | *Escherichia* | 2.3.1.47 | P12998 | 384 | 6.63 | 41463.07 |
| Biotin | *Bacillus* | 2.3.1.47 | P53556 | 389 | 6.27 | 42581.46 |
| Biotin | *Arabidopsis* | 2.3.1.47 | Q2QKD2 | 476 | 8.27 | 52171.06 |
| Biotin | *Sorghum* | 2.3.1.47 | C5WSC4 | 456 | 6.25 | 49859.11 |
| Biotin | *Wigglesworthia* | 2.6.1.62 | H6Q5U9 | 432 | 9.6 | 49033.87 |
| Biotin | *Sodalis* | 2.6.1.62 | Q2NUJ8 | 429 | 6.59 | 46891.04 |
| Biotin | *Escherichia* | 2.6.1.62 | P12995 | 429 | 5.53 | 47335.59 |
| Biotin | *Bacillus* | 2.6.1.- | P53555 | 448 | 5.43 | 50111.55 |
| Biotin | *Arabidopsis* | 6.3.3.3; 2.6.1.62 | B0F481 | 833 | 5.77 | 89031.87 |
| Biotin | *Sorghum* | 2.6.1.62 | C5WTT1 | 825 | 6.06 | 90402.01 |
| Biotin | *Saccharomyces* | 2.6.1.62 | P50277 | 480 | 5.98 | 53708.69 |
| Biotin | *Wigglesworthia* | 6.3.3.3 | H6Q5U5 | 226 | 9.81 | 25710.48 |
| Biotin | *Sodalis* | 6.3.3.3 | Q2NUJ4 | 228 | 5.92 | 24131.76 |
| Biotin | *Sodalis* | 6.3.3.3 | Q2NSY4 | 361 | 7.71 | 39000.92 |
| Biotin | *Escherichia* | 6.3.3.3 | P13000 | 225 | 5.56 | 24008.4 |
| Biotin | *Escherichia* | 6.3.3.3 | P0A6E9 | 231 | 6.11 | 24980.98 |
| Biotin | *Bacillus* | 6.3.3.3 | P53558 | 231 | 6.19 | 25160.18 |
| Biotin | *Saccharomyces* | 6.3.3.3 | P53630 | 237 | 5.01 | 26256.86 |
| Biotin | *Wigglesworthia* | 2.8.1.6 | H6Q5U8 | 333 | 9.67 | 37871.2 |
| Biotin | *Sodalis* | 2.8.1.6 | Q2NUJ7 | 345 | 6.19 | 38358.74 |
| Biotin | *Escherichia* | 2.8.1.6 | P12996 | 346 | 5.32 | 38648.09 |
| Biotin | *Bacillus* | 2.8.1.6 | P53557 | 335 | 5.63 | 36927.15 |
| Biotin | *Arabidopsis* | 2.8.1.6 | P54967 | 378 | 6.52 | 41681.52 |
| Biotin | *Saccharomyces* | 2.8.1.6 | P32451 | 375 | 8.54 | 40053.94 |
| Biotin | *Glossina* | 6.3.4.15 | TMP012922 | 1141 | 8.45 | 128366.6 |

| Pathway | Organism | Enzyme EC | Sequence ID | Length | Isoelectric Point | Molecular weight |
|---|---|---|---|---|---|---|
| Biotin | *Glossina* | 6.3.4.15 | TMP012924 | 173 | 4.93 | 19679.71 |
| Biotin | *Drosophilla* | 6.3.4.15 | Q9VNC3 | 1041 | 6.62 | 115924.3 |
| Biotin | *Tribolium* | 6.3.4.15 | D2A4T2 | 893 | 8.12 | 101244.8 |
| Biotin | *Wigglesworthia* | 6.3.4.15 | H6Q4M4 | 263 | 9.41 | 30464.99 |
| Biotin | *Sodalis* | 6.3.4.15 | Q2NWS5 | 319 | 6.46 | 34762.02 |
| Biotin | *Wolbachia* | 6.3.4.15 | Q73GI4 | 347 | 8.44 | 39245.97 |
| Biotin | *Escherichia* | 6.3.4.15 | P06709 | 321 | 7.76 | 35312 |
| Biotin | *Bacillus* | 6.3.4.15 | P0CI75 | 325 | 6.68 | 36182.48 |
| Biotin | *Saccharomyces* | 6.3.4.-; 6.3.4.9; 6.3.4.10; 6.3.4.11; 6.3.4.15 | P48445 | 690 | 6.45 | 76362.96 |
| Biotin | *Human* | 6.3.4.-; 6.3.4.9; 6.3.4.10; 6.3.4.11; 6.3.4.15 | P50747 | 726 | 5.4 | 80759.98 |
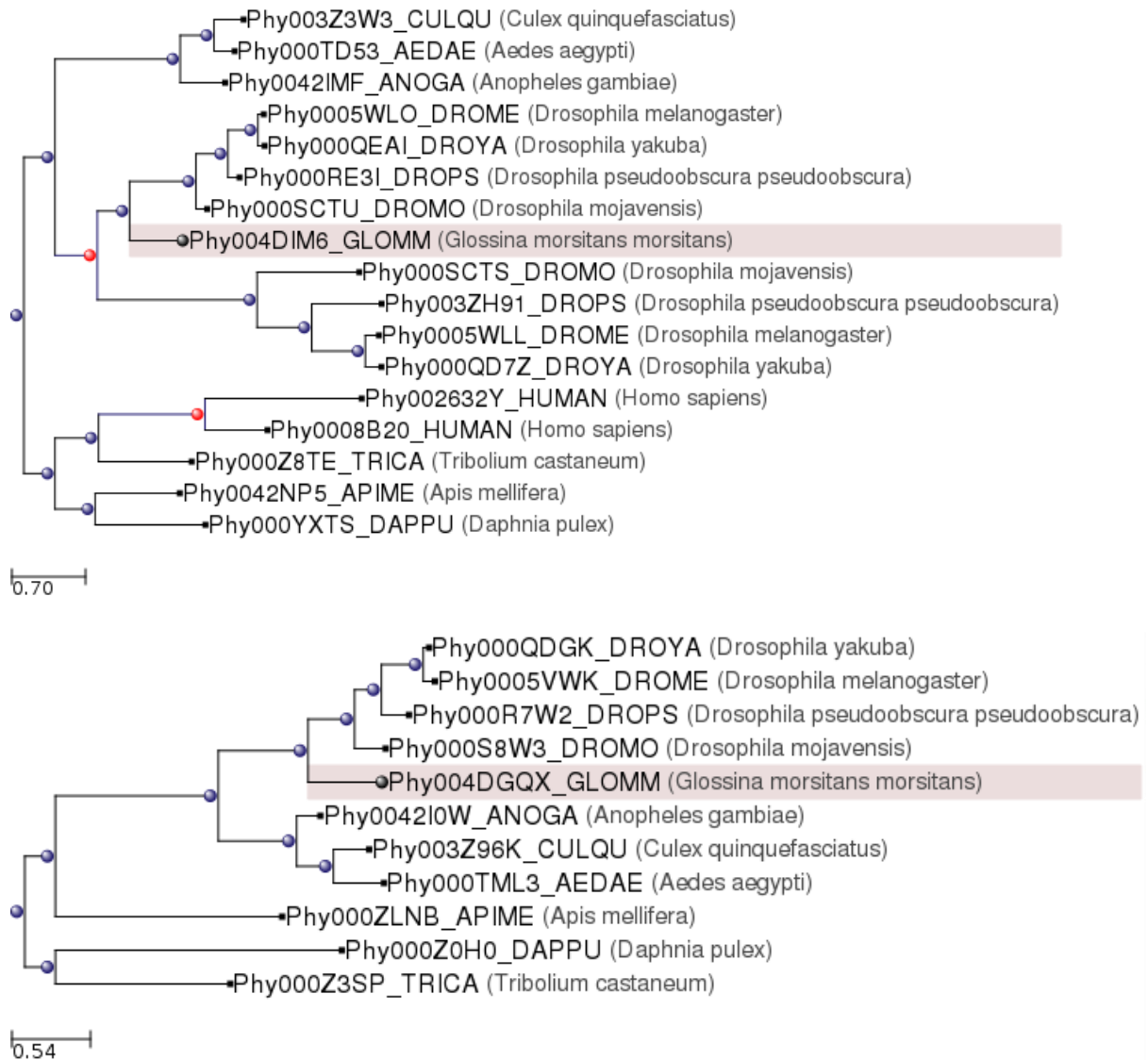


**Figure S1. Phylogenetic trees showing clustering of *Glossina* genes with close orthologs**. Pannel A represent gene GMOY009677 (Phy004DIM6) and panel B gene GMOY000849 (Phy004DGQX). Both were obtained from PhylomeDB (http://phylomedb.org)