The Establishment of Alternative Control Methods of Invasive *Eichhornia crassipes* (Water Hyacinth) Using Bioinformatics Tools

BY

EMMANUEL KUDZAI KUNDISHORA

(B1441635)

SUPERVISOR: DR. J. GERE AND DR. W. CHINGWARU

A thesis submitted to Bindura University of Science Education in partial fulfilment of the requirements for a Master of Science Education degree (Biology)

2017
NAME OF AUTHOR: KUNDISHORA EMMANUEL KUDZAI

TITLE OF PROJECT: The Establishment of Alternative Control Methods of Invasive *Eichhornia crassipes* (Water Hyacinth) Using Bioinformatics Tools

DEGREE PROGRAMME: MASTER OF SCIENCE EDUCATION DEGREE - BIOLOGY

Year this degree was granted 2017

Permission is hereby granted to Bindura University of Science Education library, to produce single copies of this project and lend such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights and neither the project nor extensive extracts from it may be printed or otherwise produced without the author’s permission.

Signed..........................................................

Date..............................................................
Approval Form

BINDURA UNIVERSITY OF SCIENCE EDUCATION

The undersigned certify that they have read and recommended to Bindura University of Science Education for acceptance of research project entitled “The Establishment of Alternative Control Methods of Invasive Eichhornia Crassipes (Water Hyacinth) Using Bioinformatics Tools” submitted by

Kundishora Emmanuel Kudzai in partial fulfillment of the requirements for a degree of Master of Science Education-Biology.

SUPERVISOR(S)

PROGRAMME COORDINATOR

EXTERNAL EXAMINER

DATE
Declaration

I, Emmanuel Kudzai Kundishora, declare that this project is my original work and affirm that it has not been submitted to this or any other University in support of any application for a degree or any other qualifications.

Signed.............................................................Date...................................................

Witness.....................................................Date........................................................

Supervisor

I, ........................................................................declare that I have supervised this thesis and am satisfied that it can be submitted to the faculty of Science Education of Bindura University of Science Education.

Date................................................................................................

Signature..................................................................................
Acknowledgement

I would like to acknowledge the continued co-operation and support of Dr. J Gere. I am greatly indebted to him for identifying the talent, ability and potential in the field of Bioinformatics as well as his guidance and encouragement throughout the study. Despite his ever busy and hectic office schedule, he offered great obligation to my work. He was available always when I needed his assistance. I would also like to acknowledge the assistance I got from Dr. W. Chingwaru who took it upon himself to take over the supervision of this project since the passing of Dr. Gere.

Special thanks goes to my family and friends, the Manicaland Basketball Association, for their unwavering support as I worked on this project. I would also like to pay mention to the Dr Chikwambi of Chinhoyi University of Technology for providing and identifying appropriate “Blast” programmes and Perl scripts.
Dedication

I dedicate this project to my wife Roselyn, my mother Mrs Betty Kundishora, My sister Barbara Mbavarira, Masimba Kundishora and Christopher Kundishora.
ABSTRACT
This study explored non-redundant National Center for Biotechnology Information (NCBI) databases to look for similarities with *Eichhornia Crassipes* (water hyacinth) transcriptome. The research was a bioinformatics data mining project which used computer programmes based on local alignment of gene sequences search principle, called BLAST2GO as well as supplementary online tools such as KEGG, KO and UniprotKB. The study focused on getting similarities of the water hyacinth transcriptome, and extrapolation of the already annotated and described genetic information to furnish the water hyacinth transcriptome in order to establish the genes and pathways that make it an invasive water weed. The project started with a water hyacinth transcriptome which was loaded as a query sequence to be “blasted” against the non-redundant gene databases maintained by NCBI. The project elucidated the involvement of genes of interest in several biological pathways using Gene Ontology and KEGG pathway analysis. Unigenes responsible for binding, cellular processes, biological processes and electron carrier were the most differentially expressed after “Blast”. Genes coding for carbon metabolism were listed as: hexokinase (HK), glucokinase (GCK), phosphoglyceraldehyde (GGP), dependent glucokinase (ADPGK ADP), phosphofructokinase (pfkc) and glucophosphate isomerase (GPI) glucokinase among others. Those coding for fast growth and water survival were listed as: mitochondrial phosphate carrier like, U1 small nuclear ribonucleoprotein 70 kDa, RNA polymerase-associated LEO1, and phosphoserine amino transferase chloroplast like. The genes for waxy cuticle formation were found to be omega-3 fatty acid chloroplast like, clathrin assembly At5g35200 and Acetyl Co A transferase like, and those for formation of air-filled roots system were found to be: 26s proteasome non-ATPase regulatory subunit 12 homolog A like and cystolic NADP-malik enzyme. These unigenes were shown to have worked in clusters to code for invasive features such as fast
growth, efficient carbon metabolism, plant defense mechanisms as well as formation of air-filled roots in the water hyacinth and, hence rendered control difficult.
# CONTENTS

Approval Form ........................................................................................................ ii

Declaration ........................................................................................................ iii

Acknowledgement ................................................................................................ iv

Dedication ........................................................................................................... v

ABSTRACT ........................................................................................................... vi

CHAPTER 1 ............................................................................................................. 1

1.0 INTRODUCTION .............................................................................................. 1

1.1. Introduction ................................................................................................... 1

1.2. Background to the study ............................................................................. 1

1.3. Statement of the Problem ........................................................................... 3

1.4. Purpose of the Study ................................................................................... 3

1.5. Objectives ...................................................................................................... 3

1.6. Research Questions ...................................................................................... 3

1.7. Significance of the Study ............................................................................ 4

1.8. Assumptions ................................................................................................... 4

1.9. Conceptual delimitations of the study ......................................................... 5

1.10. Limitations of the study ............................................................................ 5

1.11. Abbreviations and Acronyms .................................................................... 5

1.12. Operational definition of terms ................................................................. 6
1.13. Summary ................................................................................................................................. 6

CHAPTER 2 ........................................................................................................................................ 8

2.0 CONCEPTUAL FRAMEWORK AND LITERATURE REVIEW ................................................. 8

2.1. Introduction ............................................................................................................................... 8

2.2. The Water hyacinth plant physiology .................................................................................... 8

2.3. Identification of *Eichhornia crassipes* .................................................................................. 9

2.4. Dispersion .................................................................................................................................. 10

2.5. Ecology of the water hyacinths ................................................................................................. 11

2.6. The water hyacinth problems .................................................................................................. 11

2.6.1. Water transport hindrance .................................................................................................. 12

2.6.2. Environmental Problems .................................................................................................... 13

2.6.3. Threat to biodiversity .......................................................................................................... 14

2.6.4. Damage to structures .......................................................................................................... 16

2.7. Control of *Eichhornia crassipes* .......................................................................................... 16

2.7.1. The costs .............................................................................................................................. 17

2.7.2. Chemical control .................................................................................................................. 17

2.7.3. Biological control ................................................................................................................ 18

2.7.4. Mechanical control .............................................................................................................. 18

2.7.5. Reducing nutrient input ....................................................................................................... 19

2.7.6. “BLAST” ............................................................................................................................. 20
4.4. KEGG analysis results. ................................................................. 38
4.5. GO analysis. ........................................................................... 39
4.6. Biological processes score, molecular function score and cellular components distribution pie charts ........................................................................ 41
4.7. Gene lists. .............................................................................. 43
4.8. Summary. ............................................................................... 47

CHAPTER 5 ....................................................................................... 48

5.0 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS ............... 48

5.1. Introduction ............................................................................ 48
5.2. Methods and Procedures. ....................................................... 48
5.3. Major Findings ........................................................................ 48
5.4. Conclusion ............................................................................... 49
5.5. Recommendations ................................................................... 49

6.0 REFERENCES. ........................................................................... 50

APPENDIX ....................................................................................... 54

Appendix 1. KO terms list. ............................................................. 54
Appendix 2. KEGG mapper results. ............................................... 58
Appendix 3. Blast progress picture 1. ........................................... 59
Appendix 4. The water hyacinth picture. ....................................... 60
Appendix 5. Blast2Go logo. ............................................................ 62
CHAPTER 1

1.0 INTRODUCTION

1.1. Introduction

This was a data mining bioinformatics project which sought to establish the genes that code for the features that make the water hyacinth invasive. The project started with a transcriptome of the water hyacinth which was placed as a query to be compared with already published gene sequences available on the NCBI database upon which similarity means were calculated using the Basic Local Alignment Search Tool (BLAST). The computer programme or application used is a graphical user interface (GUI) based system that can be manipulated at all stages and functions to suit the researchers’ needs. It is called BLAST2GO version 3.3. The research sought to offer an alternative solution to the currently employed methods of controlling the water hyacinth. The research sought to attack the invasive characteristics at genetic and or molecular level rather than the physical, chemical and biological methods of control currently in place.

1.2. Background to the study

There is imminent upsurge in the use of Information Communication Technologies (ICTs) in every part of life and biological sciences, in particular molecular biology is not spared. Since molecular biology has lots of information in form of genes sequences sequenced from organisms, there is need to use ergonomic computer systems that can handle a lot of information with ease as compared to printed copies. Upload, search, processing and outputting of genetic information is enhanced if not made easy using computer systems such as BLAST2GO.

The water hyacinth is a fast growing and invasive water plant which has proved difficult to control using incumbent chemical, physical and biological methods. However, since there is more study
on gene ontology sequencing and mapping as well as publishing of genetic information on official state databases such as the GenBank for nucleotide sequences and Uniprot for protein sequences, maintained by NCBI, genes that make the water hyacinth invasive can therefore be matched to the already well-studied, documented and annotated genes online. This circumvents the experimentation procedure of mapping and annotation which is expensive and labour intensive. This in turn gives an idea of the pathways and processes involved in making the invasive features. Useful information such as mapping and annotation results can be collected online and can inform the manufacture of herbicides as it gives clear targets for inhibition at molecular level.

The project is a brainchild of Bioinformatics, an area within the field of Biological Sciences / Biotechnology, which relies on data mining. This novel area can be described as an overlapping bridge between Biology, Statistics and ICT. Biology provides the area of study, the organisms’ physical and genetic characteristics, and the ICTs provide the systems that can handle manipulate and process the vast information captured during biological studies in a more user friendly and rapid manner.

The databases are officially run and maintained online by the US government (NCBI) over secure networks, and it is therefore possible to upload, access and or download genetic information from anywhere via the Internet safe from malware. BLAST is a platform that matches and gives similarity mean of the query set of genes against public official databases. The genes already published online have full literature, description and annotations, such that those which come out as “hits” after search, to those known, can now be identified and online descriptions from KEGG and GO can be downloaded and transferred to the query sequence. (Wheeler et al., 2007).

The increase in costs and time consumed with respect to the control of the water hyacinth, has pushed the researcher to focus on eradicating the water weed using its genetic blueprint.
1.3. Statement of the Problem

Current control methods of the water hyacinth are proving to be costly, labour intensive as well as being less efficient. Therefore it is imperative to change focus and look at the molecular level of the water hyacinth so that genes are identified that make it invasive. The genes are then earmarked for laboratory procedures that will lead to the manufacture and design of herbicidal chemicals.

1.4. Purpose of the Study

The projects aimed to offer an alternative control method for the invasive water hyacinth. The alternative method is a molecular level based study of the genes that control the production of any features that make the water hyacinth invasive. The research seeks to provide information on the pathways involved so as to target for inhibition when making herbicides.

1.5. Objectives

The study seeks to achieve the following:

- Identify the genes that make the water hyacinth invasive.
- Map the genes that make the water hyacinth invasive.
- Extract the annotation of the genes identified from the public domain datasets.
- Establish the pathways that contribute to floating or buoyancy of the water hyacinth.
- Inform genetic engineering laboratory work on manipulation of the identified genes.

1.6. Research Questions

- What are the invasive features of the water hyacinth?
- Which genes code for the production of these invasive features?
- Which genes from the already sequenced, are they similar to?
• Where are the genes located?
• Which pathways do these genes control?

1.7. Significance of the Study

The research applied fast, accurate, efficient and ergonomic multitasking methods of search to offer a quick study of the genetic material of invasive species, so that they can be controlled at molecular genetic level. The research was deemed important as it gives a platform for the study of physical characteristics as they are linked to genes such that genetic engineering can be of use in the control of the costly water weed. This alternative removes the classical Mendelian work of actually waiting for the next generation to see the external features for study and experimentation.

The study sought to integrate the full use of computer systems to biological work and explore their strengths and limitations. The research results are useful in offering an alternative lasting method which is cheaper and effective in the long run.

The research also exhibited the potential power of ICTS in merging Biological Sciences, mathematics combined with statistical analyses systems. This in turn provides primary information for the genetic engineering techniques that can be used to manipulate the genes and inform herbicide production of target genes for inhibition.

1.8. Assumptions

The researcher assumed that the water hyacinth is difficult to control. It was also assumed that the biological information stored on the official databases such as the NCBI, were accurate, complete and up to date. It was assumed that the water hyacinth transcriptome obtained as secondary data from Chinhoyi University of Technology could be authentic source as it was from a scholarly source. Therefore it was treated as correct and accurate. It was further assumed that the computer
programme used in the study, BLAST2GO basic covered all the areas that the researcher sought to study. It was further assumed that BLAST2GO pro is an advanced user and expensive version that has little and less significant advantages over what the basic version can do.

1.9. Conceptual delimitations of the study

The study was carried on secondary data which was made available by Chinhoyi University of Technology. On the other hand, the databases used were those maintained by NCBI such that any other findings not uploaded are left out on the “BLAST” searches.

1.10. Limitations of the study

Often the “BLAST” could not run for 24 hours a day due to bad Internet connections. The researcher had to then check, at 4 hour intervals, the progress of the searches. The BLAST2GO programme would pick the database of its choice without the researchers input between Europe, Asia, South Africa and Germany. The research period was from December 2015 to August 2016, which implies any new data uploaded after August 2016 was not considered.

1.11. Abbreviations and Acronyms

ICTs Information Communication Technologies.

GUI Graphical User Interface.

GO Gene Ontology

NCBI National Centre for Biotechnology Information.

BLAST Basic Local Alignment Search Tool.

BLAST2GO BLAST to Gene Ontology.
1.12. Operational definition of terms

“Blast” running the online alignment search using Blast2Go, to find similar sequences to the query set.

Data mining is extraction of knowledge and or patterns of interest from a large amount of information and it involves intertwining multiple disciplines like database technology, statistics, machine knowledge, high-performance computing, data visualization, pattern recognition, neural networks, information retrieval, signal and image processing, including spatial data analysis.

Annotation process of picking GO terms from the GO pool attained during the Mapping stage and assigning them to the query sequences.

Mapping entail retrieving gene ontology (GO) terms related to the hits obtained after a “BLAST” search.

Hits “blast” results with significant similarity mean and estimated value.

1.13. Summary

This chapter gave an introduction as well as the background of the study by giving the gap or need that has to be filled or addressed. The purpose of the study was explored as the significance of the study. Bioinformatics is hinged on the use of ICTs in the Biological Sciences field, to ensure that data quality can be handled accurately and consistently with completeness. Operational terms were
defined and the delimitations were well elaborated. The researcher also highlighted the use of BLAST2GO Basic, a computer program in processing the primary data under study. The objectives and research questions were also stated in this chapter. These set the framework and basis upon which research is to be carried out. The chapter also looked at the assumptions of the study together with the limitations of thereof. The next chapter looks at the review of related literature to the study.
CHAPTER 2

2.0 CONCEPTUAL FRAMEWORK AND LITERATURE REVIEW

2.1. Introduction

The water hyacinth, *Eichhornia crassipes* an invasive aquatic plant is believed to have originated from Brazil and naturally extended to other parts of South America, and has extended to Africa as well (Chikwenhere, 2001, Ankney, 2012). Water hyacinth is a perennial free floating marine herb native to the South American Amazon basin (Burton, 2005). Due to its ornamental value as a beautiful flower, it was however introduced to other tropical continents as a decorative flower (Ndimele *et al*., 2011). The plant turned into a weed due to its fast growth in response to high nutrition levels in urban, industrial as well as municipal wastewaters (Trinidad *et al*., 2008). Burton (2005) notes that international experience reveals that the water hyacinth’s high adaptability, simple nutritional requirements, reproductive capacity and resistance to severe environmental conditions make it impossible to eradicate, and or difficult to control. Therefore it becomes an aggressive and invasive species, which requires urgent attention and control.

2.2. The Water hyacinth plant physiology

The water hyacinth is a plant that floats on water with round, thick glossy leaves, inflated stems, plus conspicuous lavender flowers (Trinidad *et al*., 2008). It looks very similar to the frogs bit and the water lettuce. It also has buoyant roots which possess aerenchyma cells. The mature plant has long, pendant roots, stolons and rhizomes, leaves, inflorescences as well as fruit clusters. The plants can grow up to 1 metre high but rather are usually 40cm in height. The inflorescence can bear 6 - 10 lily-shaped flowers, each with a diameter of 4 - 7cm. The stems and leaves are made up of air-filled tissue to aid the plant with its considerable buoyancy. The vegetative reproduction
is rapid under preferential conditions (Ndimele et al., 2011). Water hyacinths are the only huge water herb that are capable of floating on water without attached to the bottom FAO, 2013. This makes them highly adaptable in almost any water body sea, lake, dam or river. The water hyacinth is highly adaptable allowing it to grow, survive, reproduce and spread rapidly in freshwater. The water hyacinth can withstand extremes of nutrient supply, temperature, pH level, and also tolerate and even grow in toxic water (Ebel et al., 2006). Its growth is best in still and or slow-moving water. It is a very resilient water plant with an outstanding physiology and capable of enduring a wide range of water quality, it reproduces via more than one method. (Wang and Calderon, 2012; Julien et al., 2001).

The water hyacinth mainly reproduces asexually by vegetative means through stolons, but it also reproduces sexually by producing seeds which remain viable for 15 to 20 years (Kateregga and Sternerb, 2007). One plant under optimal conditions can give rise to 3000 plants in a time space of 50 days to cover an area of about 600 square metres in a single year (Ankney, 2012). The water hyacinth grows very rapidly such that one plant can produce nearly 140 million asexual daughter plants in a single year, which covers approximately 140 hectares. This brings in a total weight of about 28, 000 tonnes. This weed also yields plentiful seeds and its population can double in a period of 5 to 15 days (Ripley et al., 2006). The species' genetic make-up that is accountable for its reproductive methods and exponential capacity for growth is a very important contributing factor to its invasiveness (Wang and Calderon, 2012).

2.3. Identification of Eichhornia crassipes.

*Eichhornia crassipes* is an upright perennial aquatic plant that floats on water and can proliferate into dense mats, its new plantlets sprouts attaching onto the floating stolons (Malik, 2007). The roots that are always submerged have a blue-black to dark purple colour, are feathery, dense
towards the root crown, and the tips are long with dark root caps. Leaves are organized in rosettes. The leaf blades have a glossy green colour, round or largely elliptic, and sized 15 cm in width. The leaf petioles are about 30 cm in length, inflated or bulbous, and spongy especially near the bottom (Gichuki et al., 2012). The water hyacinth has inflorescence showy as a spike above the rosette and is also approximately 30 cm long. Its flowers are lavender-blue and have a yellow blotch, grow up to 5 cm wide, and to some degree 2-lipped. The flowers are comprised of 6 petals and 6 stamens. The plant bears fruit that is a 3-celled casing with numerous seeds. In total the average size ranges from 60-120 cm in length (Kateregga and Thomas, 2007).

2.4. Dispersion

There are mainly three means by which the water hyacinth can be spread from one place to another. The water hyacinth can be dispersed by water that is water-borne dispersal a method that requires a link between water bodies. The water hyacinth in this case simply flows with the water currents and proliferates to new sites (Center et al., 2002). Secondly the water hyacinth is dispersed via the diving eater-birds which eat and swallow the water plant’s seeds as they go through the mud in search for food. The seeds are then passed through the gut undigested, and get deposited to new areas where they germinate even after long periods of dormancy (Gonzalez et al., 2006). The third way is via people as they deliberately export it as an exceptionally beautiful ornamental purple flower with high demand and great price in aquariums and ornamental ponds (Burton, 2005). Another way is by cargo where rafts of the weed are hooked on the underside of the marine transport engines and get carried away to new lands. These are some of the ways in which the water hyacinth has managed to cover the whole world with little exceptions such as Congo River, Egypt and Uganda (Masifwa et al., 2001).
2.5. Ecology of the water hyacinths

The native range of the water hyacinth is Amazon basin, Brazil and surrounding locality. It however occurs and is widely spread in non-indigenous areas and states such as Alabama, Arizona, Arkansas, California, Colorado to Virginia and Washington (Ankney, 2012). *Eichhornia crassipes* is growing on 5 continents (Lowe *et al.*, 2000). This water plant species has been noted in New South Wales, Australia, in the Bahamas, in Bangladesh, Benin, British Virgin Islands, in Africa, and Asia (Fayad *et al.*, 2001), (Moorhouse *et al.*, 2001). In Africa the plant is also prominent in Zambia, and Zimbabwe (Ankney, 2012). This shows that the plant has covered almost all parts of the world due to its high adaptability. As such it has been declared class 2 plant under the Land Protection (Pest and Stock Route Management) Act 2002. A Class 2 pest is the category that has at present spread over extensive areas whose impact is so serious that there is need to try and control it as well as circumvent additional spread onto new areas that are still free of the pest. The water hyacinth easily grows on any open or sheltered water surface because of its extraordinary buoyancy (Perez *et al.*, 2010).

2.6. The water hyacinth problems

Water hyacinths can proliferate into dense mats so thick that a person can walk on them. When this happens, water hyacinths hold and block rivers and canals, thereby causing less water flow. The mats can become so dense to form an herbivorous barrage leading to damaging and dangerous flooding. The mats also greatly affect transport engines as well as water reticulation centres. The water weed is also habitat to parasites such as mosquitoes and also reduce the oxygen demand in water (Ndimele *et al.*, 2011). As such, it is regarded a serious threat to biodiversity. There are several control methods which have been employed to manage and control the weed but most if not all have proved to be either less efficient and or costly in the long run. Problems associated
with *E. crassipes* are commonly a result of its prompt growth rate, its capacity to successfully compete with other aquatic plants, and its ease of propagation. Water hyacinths have potential to cause a lot of problems when its mat proliferates rapidly and covers vast areas of fresh water. Some common problems are discussed here.

2.6.1. **Water transport hindrance.**

Hindrance to water transport is a global problem posed by blocking water hyacinth mats (Gichuki *et al.*, 2012). Burton (2005) states that harbours and docking areas can be inaccessible due to hindrance mats of water hyacinth, which clog and choke marine engines. Freshwater rivers and canals can also become impassable as the densely intertwined carpets of the weed disturb the small boats and canoes (Nimele *et al.*, 2011). On Lake Victoria it is becoming a serious hazard to lake transport as large floating islands of water hyacinth emerge, while most inland waterways of South East Asia have been almost disused and abandoned (FAO, 2013). Water transport engines are greatly affected by debris in water and the thick mats caused by water hyacinths not only slow down but at times damage the engines and stop them totally. It is however cumbersome and expensive to keep cutting and shoving the water hyacinth mats off the waterways (Villamagna and Murphy, 2010).

On a global scale water hyacinth mats reduce maneuverability in water transport systems such as boats and ships. The infestation of *E. crassipes* resulted in increased transportation costs, blockage of irrigation canals, and difficulties in electricity and water extraction. As a result of *E. crassipes* infestation, the Kenyan authority of port Kisumu recorded a 70% decline in economic activities. This species can negatively affect recreational fishing when it makes it difficult to access fishing dregs and inhibiting boating (Gichuki *et al.*, 2012).
2.6.2. Environmental Problems

Environmentally *Eichhornia crassipes* grows rapidly to form thick mats which decrease the light and oxygen availability in water (Ivanov, 2006). Field experiments which involved removing *E. crassipes* from Lake Victoria by mechanical means (cutting), showed that there was a substantial upsurge in dissolved oxygen in the lake (Osumo, 2001).

For Africa, *Eichhornia crassipes* can negatively impact water availability because of its high evapotranspiration rate as well as capability to take advantage of limited water reserves, this consequentially results in a yearly 7 billion loss of water from River Nile (Julien et al., 2001). In Asia, China in particular, *E. crassipes* has aggravated water pollution as it absorbs heavy metals and discharging them at death (Ebel et al., 2006), Jianqing et al., 2001). It is noted that in the past decade the water hyacinth rapidly spread throughout many parts of Africa leading to great concern.

Thick floating rafts of water hyacinth which float on the water’s surface, restrict light penetration to the complete elimination of other native plants, as well as reducing the air interchange between the water’s surface and the atmosphere (Malik, 2007).

Decay of *E. crassipes* can generate anoxic (low oxygen) conditions in water. Water-hyacinth mats increase flooding in rivers and canals as they form dams. When the *E. crassipes* was removed from Lake Victoria by cutting, the resultant was a substantial upsurge in dissolved oxygen (Osumo, 2001).

Locally in the region and in Zimbabwe, the *E crassipes* has duplicated and extended its problems in inland waterbodies, and has been reported to have disturbed fishing and water works activities in lake Chivero of Harare (Chikwenhere, 2001).
General public often point out issues of localised water quality decline. This becomes of substantial worry where people come to fetch water and to wash (Osumo, 2001). However to quantify this problem of quality at a mathematical continuum is often extremely difficult. This species has been reported to reduce water availability in Lake Victoria Basin, which led to social conflicts over the absence of clean portable water (Mailu, 2001).

*E. crassipes* has great potential for high socio-economic effects if they are to be introduced to the Great Lakes. It can pose risk to human health as it provides an ideal habitat for mosquitos, and this may proliferate the risk of mosquito-borne diseases (Jianqing et al., 2001, Mailu, 2001). Dense mats of the water hyacinths trap sediments, reduce the flow of water in irrigation channels and waterways thus can provide ideal breeding environment for mosquitoes. The sediment levels rise with increasing *E. crassipes* profusion (Wang and Calderon, 2012).

### 2.6.3. Threat to biodiversity

In areas where water hyacinth is abundant, other marine plants find it difficult to thrive. Its competitive advantage causes a disparity of imbalance at the aquatic micro-ecosystem. This implies that a range of fauna which normally depends on a diversity of plant life for its survival will be driven into extinction. Fish diversity is also provoked with some, however benefiting but others suffering from the propagation of water hyacinth. Algae, which makes up a major part of the basis of the marine food chain, can be covered out by dense mats of water hyacinth. This results in the decline in algae population, and this in turn disrupts the whole food web in a water body (Coetze et al., 2006).

*Eichhornia crassipes* can form impenetrable single species stands that often do not avail optimal habitat environments and or food to native wildlife, and thus may limit entry to and contact with
the water for some species. The native wildlife populaces may be pushed to relocate or die, hence this ultimately results in a loss of biodiversity and an interference in the equilibrium of the ecosystem. Water hyacinth causes physical obstruction of water ways, alters the food web, lowers overall pH, and out-competes more sensitive native species (Masifwa et al., 2001).

The effective invasion of invasive water hyacinth plants is sometimes attributed to the absence of its natural predators and competitors. In its native habitat, the water hyacinth is food for manatees (Trichechus manatus) which keep their populations in almost constant equilibrium (Center et al., 2002). The lack of any equally predatory and competitive aquatic fauna in other locations notably promotes the successful proliferation of the water hyacinth’s populations in most rivers.

*Eichhornia crassipes* can possibly modify predator-prey relationships to its survival advantage. As sited in San Joaquin Delta, California, insect densities where lesser in patches of *E. crassipes* and there was a remarkable variance in insect pest composition between *E. crassipes* and the native pennywort (*Hydrocotyle umbellata*) (Jimenez, Balandra, 2007).

*Eichhornia crassipes* is an extremely competitive plant is capable of rapidly growing, reproducing and spreading. *E. crassipes* dislodges native species successfully, reduces biodiversity significantly, limit recreation, lessen aesthetic value, and decrease water quality and flow.
2.6.4. Damage to structures

During flooding, the water hyacinth rafts of plant material collect and deposit up at fences and bridges and this in turn collect other floating debris. The resultant combined weight can outweigh the structures and hence they collapse. Some landing sites have been totally shut down and are no longer functional. For example, Katosi landing site in Mukono District closed down because of the overwhelming mats of the water hyacinth. Fishermen and porters relocated to less affected landing sites like Casey in Mpigi District and others in Jinja District in Uganda (Kateregga and Sternerb, 2007). The other implication of the water hyacinth plague to the fishermen is physical obstruction of entrance canals to fish landings. Masses of the weed mats that are swept by storms and water currents block these water channels. The fishermen are thus forced to push the boats in between the intertwined mats of weed. They spend more than 30 minutes forcefully shoving nine-metre 16 fibreglass canoes over a distance of 100m. This is tedious and laborious. For fishermen who operate basket traps to catch tilapia fish, lungfish in marginal swamps, and others who set gill nets are greatly affected by large rafts of the water hyacinth. The obstruction hence leads to loss of the set fishing gear and prime fishing spots (Villamagna et al., 2010).

2.7. Control of Eichhornia crassipes

Globally there is a general consensus among environmental scientists and managers that there is no single and totally effective means to eliminate *E. crassipes* this points out the best option as the integrated management and control of the herb (Jiménez and Balandra, 2007). Control measures that have been proposed and put into practice include the chemical control, physical control and biological control. All the three have been combined into an integrated control methods. Zimbabwe has been engaged in the biological control of pests, and floating aquatic weeds, dating back to
1950s. The country has been at work with international collaborators as a way of attaining sustainable biological control technologies (Chikwenhere, 2001).

### 2.7.1. The costs

The costs associated with the management of water hyacinth worldwide are so huge for the economy and ecology, such that there is imminent need for environmentally safe and economically sustainable control measures which can provide lasting solutions to invasive weed infestation. Water hyacinth is very extraordinary and vigorous budding angiosperm plant belonging to the fresh water aquatic habitat of the family Pontederiaceae (Ankney, 2012). The water hyacinth is usually referred to as the world’s vilest aquatic weed, also called an economic drain because of its ability to cover whole waterways prompting funds to be channelled towards its control.

It is incumbent therefore to look into its genetic make-up and try to manipulate the genes that code for the physiological features and biochemical pathways that make the water hyacinth invasive and highly adaptable. The features include the buoyancy in stems and leaves, bulbous petiole as well as the aerenchyma root cells and also the biochemical pathways that control and support the floating mechanisms and physiology.

### 2.7.2. Chemical control

Water Hyacinths are tough and almost impossible to eradicate. Arsenic was employed in the US on a large scale and it only partially cleared these water weeds and on the other hand poisoned the ecosystem, threatening biodiversity (Ripley et al., 2006). Explosives and fire were also used in an attempt to clear the weeds, but since the plants are capable of rapidly reproducing and colonizing any available water body, the smallest fragments left behind simply grew back into new colonies.
The use of herbicides to control water hyacinth has been done for many years. The commonly used herbicides include 2.4-D diquat and glyphosate. It has been established that there are good results when these herbicides are administered over small infestations but less accomplishment with larger areas. The herbicides can be applied from the ground or from the sky and skilled personnel is required to operate. When herbicides are used, there are environmental and health related issues, particularly when people collect water for washing, drinking and cooking (Jiménez and Balandra, 2007).

2.7.3. Biological control

Two weevils, a moth and two types of fungi have been introduced to successfully control the plant. These reduce the proliferation considerably and in turn are not harmful to the water organisms. Other creatures that keep the plant in check include fish [Chinese grass carp (Ctenopharyngo idella) and Tilapia melanopleura and T. mossambica)] and manatees. Bioherbicides developed United States Florida’s to control E. crassipes infestation, and the pathogen Cercospora piaropi, the Myrothecium roridum (fungus) as well as the water hyacinth-specific pathogen, Alternaria eichhornia (Coetze et al., 2006). Research in the field show that the use of C. piaropiare and the surfactant Silwet L-77 were positive in decreasing the biomass of E. crassipes Coetze, (Byrne and Hill, 2006).

2.7.4. Mechanical control

Physically removing the water hyacinth mats is considered as the best short-term remedy so far to the propagation of the water plant. However, it is costly, whether using land-based clamshell, draglines, booms, bucket cranes, or otherwise, water based equipment like mowers, dredges, barges and or other specifically crafted machinery for aquatic weed harvesting. These methods are
appropriate and limited to only comparatively small areas. A lot of these techniques need the backup support of a taskforce of water and land-based vehicles for shunting of the huge amounts of water hyacinth debris which is removed. Mats of water hyacinth can be enormous, very heavy, and can have a density of up to 200 tonnes per acre (Burton, 2005). Manual removal of water hyacinth is appropriate only for particularly small areas, since the plant can regenerate and grow again within a short space of time. It is not only problematic, but also labour intensive, rigorous, and above all in some areas serious health risks associated with this work include crocodiles, hippopotamus and bilharzia in Lake Victoria as examples. Additionally, the transportation of the reaped weed is again costly, since it has such a high water content, and hence becomes heavier, and this adds to the total cost of fuel required to complete the whole removal process. Chopping can however decrease the volume and the water content considerably (FAO 2013).

2.7.5. Reducing nutrient input

Aside the mainstream forms of control Harley, Julien and Wright (2001) point out another way, which is the reduction of nutritional elements required by the plant to the water. Although this can be regarded as a preventative method, it can on the other hand be claimed that a reduction in basic nutrients in the water body results in a decreased proliferation of water hyacinth. Over the past few decades there has been a significant increase in the level of nutrients dumped into waterways from industrial and domestic sources as well as from land where chemical based fertilisers are used, or where clearance has caused an increase in surface run off (Julien et al., 2001). This in turn increased proliferation rates of the water hyacinth due to its ability to feed on and utilise the traces of food and nutrients found in waste material. The most promising control measures proved to be the biological controls, with hundreds of them having been studied and documented (Center et al., 2002).
2.7.6. “BLAST”

Several search engine algorithms have been developed to search online databases for sequences which are similar to a query sequence, the one that contains the unknown sequences of interest to be studied.

The most outstanding search algorithms used in searching online databases since 2003 are those built on “BLAST” (Basic Local Alignment Search Tool). These have high efficiency when working with protein coding sequences (Buffalo, 2015).

“BLAST” accelerates the search task by looking up for short regions of exact match between the query set and the online database sequences and then scrutinise the sequence that connects these regions to see if there is a prolonged stretch that perfectly matches. “BLAST” uses a query sequence obtained by arbitrary screening of cDNA, to equate it against a GenBank online database maintained by the NCBI. So BLAST helps us by giving information such as is our sequence similar or dissimilar to the already well-studied and published ones. “Blast” then picks the annotations and descriptions stored as GO terms on the websites for those hits that have matched the sequences on the query set (Shevnin et al., 2013).

The analysis will be run on a super computer maintained by NCBI which is the server and the personal computer running “BLAST” becomes the client. So there is a client-server relationship created online in which the researcher’s computer will link and access information contained in the online sever to look for matches. “Blast” x is an available option on BLAST2GO basic which allows the client to send a query sequence to the server and the sever responds by giving hits of similar sequences. The hits are those sequences that resemble the online datasets (McGinnis and Madden, 2004).
2.8. Summary

Water hyacinth is a floating monocot flowering in the family Pontederiaceae. It has fibrous roots which can grow to a meter in length. It has stolon type stems mostly. It is a very aggressively reproducing plant with sympodial rhizome creeping in mud. At the terminal end of every sympodial branch is a rosette of broad spoon – shaped leaves that have very turbinate swollen petioles. Petioles are commonly bulbous, spongy and at times can non–bulbous. The water hyacinth leaves are glossy, smooth and bright green in colour with rusty yellow margins. Budding from the core of the rosette emerges a sheathed scape about 6 – 12 inches tall of violet blue, cup shaped flowers. These flowers are bisexual, zygomorphic, conspicuously gamophyllus studded with 6 stamens. It primarily reproduces asexually via stolons, which form offspring plants. It also bears 3 – celled aggregate fruits and ribbed seeds. Each plant yields thousands of seeds each year that can remain dormant but still viable up to 20 years. This plant can proliferate 300 plants in a short period of time if nutritional requirements are availed especially via sewage and industrial wastes. Its tolerance of a wide range of extreme conditions helps this plant to become aggressive and invasive covering almost all parts of the world. The plant has become a pest and many control methods have been employed from mechanical, chemical and biological. However the success in controlling the water hyacinth has faced several challenges and this prompted the need to look into the genetic make-up and try to control it from within.
CHAPTER 3

3.0 RESEARCH METHODOLOGY

3.1. Introduction

This chapter outlines the research methods and methodology used to conduct the study. The chapter also highlights the paradigm that influenced how this particular research was to be done. In this chapter the computer programmes used are highlighted, how they were installed and run. The data worked on was obtained from Chinhoyi University of Technology Department of Biotechnology. The data were aligned online using a “BLAST” programme and were subjected to mathematical analyses that included calculating similarity means as well as correlation with the non-redundant databases. Since this was a data mining project, it was entirely carried out online on a computer.

3.2. Research objectives

Since petabytes of useful data sit on mega computers and servers around the world, key insights into biological questions are stockpiled not just in the unanalysed experimental data in the Personal Computer hard drives, but rather are spinning around disks in a data centre called NCBI, which is situated thousands of kilometres away. However with the use of the Internet, that data can be accessed, downloaded and or updated remotely using a connected computer.

The objectives of this study therefore were to:

- Identify the genes associated with the morphological features of the water hyacinth.
- List the genes that code for specific pathways involved in making the water hyacinth invasive.
- Identify the genes and pathways associated with floating.
- Identify genes and pathways associated with formation of fibrous roots.
- Identify genes and pathways associated with nutrition mechanisms of the water hyacinth.
- Recommend the manufacture of herbicides using the identified pathways as targets for inhibition.

3.3. Research Design

Influenced by the exploratory paradigm, the research design employed is the descriptive exploratory design. A set of sequences, the transcriptome was used as a query set in “BLAST”. “Blast” was run on default settings to obtain hits. Hits are the results which are a collection of similar sequences to those available on GenBank. Mapping and annotation was done on the hits to transfer the known descriptions of the already studied sequences to the ones that match them on our unknown query set. “BLAST” helps give more information on the query set by downloading the data stored about the segments of genes that are aligned to the query set.

3.4. “BLAST”

Many search engine algorithms have been developed to search online databases for sequences which are similar to a query sequence, the one that contains the unknown sequences of interest.

The most important search algorithms useful in searching online databases since 2003 are the group of algorithms built on “BLAST” (Basic Local Alignment Search Tool). These have high throughput efficiency when working with protein coding sequences (Buffalo, 2015).

“BLAST” speeds up the tedious search task by looking for short regions of perfect match between the query set and the database sequences and then scrutinise the sequence that connects these regions to see if there is a lengthy stretch that perfectly matches. “BLAST” uses a query sequence
obtained by arbitrary screening of cDNA, to compare it against a GenBank online database maintained by the NCBI. So “BLAST” helps us by giving information such as is our sequence similar or not to the already well-studied and published ones. “Blast” then picks the annotations and descriptions stored as GO terms on the websites for those hits that have matched the sequences on the query set (Conesa et al., 2005).

The analysis was run on a super computer maintained by NCBI, which becomes the server, and the personal computer used by the researcher, becomes the client. So this is a client-server relationship created online. “Blast” is an available option on BLAST2GO basic which allows the client to send a query sequence to the server and the sever responds by giving hits of similar sequences. The hits are those sequences that resemble the online datasets.

### 3.4.1. Sampling

The whole transcriptome was treated as the sample, so purposeful sampling was preferred to random or stratified sampling techniques. In other words the whole transcriptome represented the sample.

### 3.4.2. Population

The whole transcriptome of the water hyacinth was used as the sample to be compared to the whole population of published gene sequences contained in the online databases such GenBank, Uniprot/Swissprot maintained by NCBI. The NCBI databases contains around 26 000 non-redundant gene sequences of varying base lengths from 600 to 1200 kilo base pairs.
3.4.3. Pilot Study

In order to be familiar with the computer application software to be used, the researcher went through rigorous practise using various “Blast” based search engines. At some point manipulating the source code, in the command line changing the programming language in order to tailor the application so that it suits the researcher’s needs. Example sequences were readily available on the NCBI and UniprotKB websites for use during practise. BLAST2GO was seen fit and was chosen for this study.

3.5. Internal and external validity

To ensure internal validity which is the ability of research instruments to carry out research efficiently, the computer application software to be used were tried and tested during pilot study. Adjustments were made in command prompt mode to change the source code to redirect the “Blast” runs. The precision of Blast2GO was tested with a dataset for which functional annotation information was available. The methodology and results of this evaluation are provided as additional material that are accessible on the B2G site (Wheeler et al., 2007).

External validity was ensured by using an authentic NCBI data base run by the government of the United States of America, and Uniprot/ Swissprot, such that any results found therefore can be reproduced in any other research by another researcher.

3.6. Requirements

A fast computer with processing speed of above 3 Gigahertz. For this study an Intel inside core i7 v Pro CPU was used.
Since the computer used has Windows 10 operating system installed, the BLAST2GO version downloaded and installed was the windows version with architecture of 64 bit.

The research required uninterrupted or unlimited Internet connectivity and for this study, the researcher used TelOne’s ADSL unlimited Platinum which was availed as WIFI.

For statistical analyses BLAST2GO offers inbuilt R statistic packages which are displayed to the user via user friendly GUI menus icons and pictures to easily produce high quality graphs and pie charts in colour and in 3d.

Google drive and Dropbox were used as cloud network backups in case any data on hardware is tampered with or lost accidentally.

A user manual for BLAST2GO available on the BLAST2GO website was also downloaded and printed for quick reference and guidance.

3.7. Procedure

![Diagram of BLAST, Mapping, and Annotation process]

**Figure 1.** “Blast” procedure
The basic steps in “blast” are illustrated in the diagram above where a query set of sequences was loaded onto “BLAST” program and “BLAST” was run, after “blast” was run, came mapping stage then annotation stage.

When Blast2go was downloaded and installed, the query sequence in this case the water hyacinth transcriptome, was loaded on to BLAST2GO. Loading of a transcriptome takes 20 seconds. Once loaded, the next icon labelled “blast” was expanded which has options such as cloud “blast” which sends the query set direct to the server, NCBI “blast”, Amazon “blast” and local “blast” for downloading data from the FTP sites and run “blast” on your PC. NCBI “Blast”, the default was chosen and “blast” was run against the non-redundant database, using Blastx-fast option, with the “blast” expectation value (E-value) set at 1.0e-3 which is the default plus the “blast” descriptor annotator set on.

“Blast” results were collected into one folder and labelled “Blast” and the mapping was done to pick the GO terms associated with the hits. Mapping provided important clues about which chromosome contains which genes and also precisely where the genes are located on chromosomes among the hits. Interpro scan an optional stage was also done for protein function classification.

After mapping, annotation was done to furnish biological pathways information to the sequences. The mapping and annotation were done under default settings to ensure maximum searches and reach out. However there are options to change the default parameters for advanced users. These options are available in other varieties of blast based algorithms.

After annotation the next stage involved analysis.
3.8. Analysis

The first option under analysis tools was GO-Slim which has Enzyme Code and KEGG the second option was Enrichment analysis which used the Fischer’s exact test. Blast2Go application makes possible the direct statistical analysis of gene function data. One major analysis option is the statistical calculation of GO term enrichments in a cluster of interesting genes if matched to a reference group. This option was introduced in Blast2Go via the integration of Gossip. Gossip is another programme that calculates Fisher’s Exact Test making use of robust False Discovery Rate (FDR) correction for compound testing and then gives back listed significant GO terms presented in the order of their corrected or one-test $P$-values. Moreover Blast2GO has a number of statistical charts concluding the results obtained during “blasting”, mapping or annotation. Bar graphs, and or pie charts of annotation statistics (GOs/Seqs), similarity / $E$-value distributions, and EC distribution profiles can be produced, saved and or printed (Bluthgen et al., 2004).

Also available in “BLAST” are statistical packages such as R statistics used to make graphs. The statistical package R could be loaded with statistical formulae via command line to change the software to suit this research. Graphs were chosen at different levels to give the desired detail level 3 shows low detail and the higher the number for example 10 would give much more detail on cellular function and biological processes. The same embedded statistical software also drew charts and the researcher chose high definition 3 d charts after analysis.
Figure 2. Application data flow diagram.

The diagram above show how Blats2Go is run. Starting from left to right there is:

1. “Blast”: in which a set of selected query sequences is searched (“blasted”) against either the non-redundant NCBI or other databases.


3. Annotation: sequences get furnished with biological information based on an annotation rule which considers parameters set by the researcher (usually set to default).

4. Statistical analysis: option to analyse GO term distributions, correlation, regression and differences between groups of sequences.
(5) Visualization: results of annotation and statistics can be visualized on the GO DAG and at every stage, charts and graphs are obtainable for ongoing evaluation of the progress of analysis (Ashburner, 2000). Data can also be saved and or exported or transformed into different formats. http://bioinformatics.oxfordjournals.org

Using KEGG mapper was another option done by the researcher to triangulate the methods to ensure quality results. A comparison was then made with the basic BLAST2GO versus KEGG analysis. Kegg mapper was accessed online and KO lists were uploaded and the Kegg pathway analysis was done. Kegg colour search pathways provided graphical information on which pathways are represented by the transcriptome (Kanehisa and Goto, 2000).

3.9. Summary

Blast2GO an all in one bioinformatics computer software package was useful in functional annotation of (novel) sequences and its subsequent analysis against annotation data contained on online databases. The main function of BLAST2GO is to ascribe as much information about the biological function of gene or protein sequences by looking up on the diverse public online resources like comparison algorithms and databases.

The query sequence was loaded onto BLAST2GO and it discovered sections of resemblance between biological sequences (the query set versus the online database). The program compared nucleotide or protein sequences to sequence databases and calculated the statistical significance given as the e value on the hits. This application software recognised already characterized identical sequences, and transferred its functional descriptions and labels to the (novel) uncharacterized sequences.
Basically we were moving from what we don’t know then match it with what is known already and then extrapolate meaning from the known sequences and then attach meaning to the unknown.
CHAPTER 4

4.0 RESULTS ANALYSIS AND DISCUSSION

4.1. Introduction

About 27000 non redundant online gene sequences were explored to search for similarities with those on the query set of the water hyacinth transcriptome. However to streamline the search best hits and unique hits lists were also obtained by assigning Perl scripts to run best hit and also unique gene sequences. Best hit list was produced by a Perl script, a computer programming code obtained from Dr Chikwambi of Chinhoyi University of technology (Department of Biological Sciences), to remain with those gene sets that had higher significant e values. Thus leads to a decrease in the amount of “actual” gene to a lesser value by removing duplicated genes.

Data obtained in form of gene lists were usable in the mapping and annotation in KEGG orthology terms to establish the pathways and genes linked to the invasive features of the water hyacinth.

This chapter therefore seeks to present, analyse and discuss the research results.

The purpose of the study was to find similarities between the query set of the water hyacinth transcriptome and the published genetic databases by exploring the online NCBI databases such as GenBank (of nucleotide sequences) and Swissprot (of protein sequences).

4.2. Blast results

Data distribution of Blast2Go result against NCBI non-redundant databases.
Figure 3. Sequence distribution.

Blast2Go results of data distribution bar graph shows that of the total non-redundant sequences, 19,343 were annotated, 10,128 were mapped, 25,235 had “blast” results, 612 did not have “blast” hits and 419 did not have any “blast” result. Those without any “blast” hit are useful to indicate curation status to signify gene products for which no significant information could be extracted. The results depict the efficacy of Blast2go in retrieving similarity information of the query set against the non-redundant databases. A few sequences had no “blast” results which means that at the time of “blast” search the sequences are either not well represented or that the transcriptome contains some non-coding portions. The results further suggest that the “blast” hits results where about 25,235 out of 26,678, those mapped dropped to 10,128. This can be explained by the fact that some gene sequences are representing the same position or locus hence the large overlap. However the figure shot back after annotation to 19,343 suggesting that there is difference in their function as annotation provides the finer details of gene function, process or localization.
Data distribution of Local “Blast” results against downloaded reviewed uniprot *Viridi plantae*.

**Figure 4.** “Blast” results against reviewed *Viridi plantae*.

“Blast” output of the genes “blasted” against the reviewed UniProt *Viridi plantae* dataset, a total of 26,680 genes had hits with varying similarity means. 26,467 of the total hits had the best hits and 26,333 were the unique ones. The downloaded FTP file contained *Viridi plantae* gene sequences which are reviewed meaning that they have been well revised and updated. However only 13% had no “blast” hits a small figure compared to the total gene set. Since the reviewed uniprot *Viridi plantae* dataset showed a slight difference between the best hits, unique hits as well as the total hits, it implies that there were a few overlaps. This can be attributed to the fact that we are looking at a *plantae* data set as compared to the open dataset of all organisms used in Blast2go. Not surprising is the fact that some hits could however be obtained with similarity to non-plant organisms such as rats and human beings. Proof that all organisms share the same amino acids that
make up their genetic material. Data distribution of Local “Blast” results against downloaded unreviewed uniprot *Viridi plantae*.

![Pie chart showing data distribution](image)

**Figure 5.** “Blast” results against unreviewed *Viridi plantae*.

The unreviewed uniprot *Viridi plantae* dataset contains those sequences not yet fully revise and updated but however there were some interesting results since from the total of 34,873, 26,950 had best “blast” hits of which 26,607 were unique. This shows a slight difference between the total hits, best and unique hits, as compared to the reviewed data set. However the values are within the same reasonable range which suggest that there is little information left out in the unreviewed dataset.
Figure 6. “Blast” hits for combined reviewed and unreviewed Viridi plantae.

The “blast” output combined (reviewed and unreviewed uniprot) had 53836 total hits whose best were 27 168 and the unique being 26 826. The combined results of the blasts combined show that their unique sequences which represent the actual genes not duplicated are averaging to the unreviewed and the reviewed data set. An indication that the unreviewed data set contains little information on the genes of the water hyacinth.

4.3. Comparing genes of the Water hyacinth to other model organisms.

Categories of genes similar to other organisms.
Figure 7. “Blast” hits compared with other organisms.

The gene annotation visualisation shows that the water hyacinth genetic material has great similarity with the plants *Arabidopsis thaliana* and *Oryza sativa*, and also has a close similarity with the unknown sequences denoted other in the graph. Further research is therefore imminent to establish this outstanding similarity with an unannotated organism.

Figure 8. Comparison of unigenes categories and pathways of water hyacinth to other related organisms.
Representation of carbon metabolic pathways such as the citric acid cycle shows that the water hyacinth is probably a C4 plant, and also presence of CAM metabolic pathways indicate that under certain conditions this plant can carry out CAM processes. This gives clue to the plant’s ability to strive in almost any water body. This evidence also supports the plant’s high nutritive efficiency in converting whatever available material such as waste to become its fertilizer including toxins.

Lipid metabolism well represented in the metabolic pathways also supports the idea of waxy cuticles on the leaves which make the plant better adaptable to water life.

4.4. **KEGG analysis results.**

![KEGG pathways](image)

**Figure 9.** Pathways represented in KO terms for reviewed uniprotKB viridi plantae.

KEGG mapper obtained the following results on pathways of interest that confer to invasive features of the water hyacinth which are buoyancy, efficient carbon metabolism represented by 1021 unigenes, waxy cuticles on leaves linked to fatty acid metabolism represented by 1032 unigenes as well as fast growth which is related to ATP synthesis efficiency which had 782 unigenes. The water hyacinth also showed a considerable efficiency in carbon fixation as elucidated by 678 unigenes aligned to carbon fixation. Transamimation is also represented by a
lower value of 236 unigenes. The efficiency of the water hyacinth in metabolism is the key aspect of its success in invasiveness. However, 1290 KO terms were not found, this is either the terms are not yet fully updated online or there might have been errors during KEGG analysis. Further analysis using other tools, to explore the KO terms not found is recommended.

4.5. **GO analysis.**

Colour search pathway produced the following results:

*Arabidopsis thaliana*

*Oryza sativa* (rice)

*Saccharomyces cerevisiae* (yeast)

EC code distribution

GO level distribution
The GO level distribution of sequences of the Water hyacinth show that the biological pathways responsible for making the features that confer to invasiveness of the water weed are well represented in the transcriptome. The Water hyacinth possesses intricate biological pathways which are coded for and controlled by gene sets embedded in its transcriptome. Of the biggest value are the structural molecule activity, cellular biosynthesis as well as translation. About 23% of unigenes and 18% were informative. Sequence BlastX analysis revealed that there were numerous adenosylhomocysteinase-like proteins, non-membrane bound organelles, biological regulation, transporters and cellular biosynthetic processes in *E. crassipes*.

**Figure 10.** Differential GO terms of the water hyacinth against NCBI databases.
4.6. Biological processes score, molecular function score and cellular components distribution pie charts.

**Figure 11. Biological Processes Score.**
Pie charts of Orthologous Group scores present that the most genes were involved in coenzyme transportation and metabolism, cell and organelle, structural molecule activity as well as growth and reproduction. Gene Ontology (GO) analysis groupings confirmed that unigenes responsible
for binding, cellular process, biological processes and electron carrier were the most differentially expressed unigenes for the blast between the water hyacinth transcriptome and online database libraries. All the collective results suggest that *E. crassipes* can survive in any water condition by efficiently controlling as well as regulating transporter gene expressions, certain metabolism processes, coupled with specific signal transduction pathways, binding, enzyme regulators and cytoskeletal construction.

### 4.7. Gene lists.

Gene lists generated from blast2go results to show relative distribution of hits amongst biological processes, molecular function and cellular component. The following gene listed in the table show the relative length of sequences, their e value as well as the enzyme code for the water hyacinth transcriptome.

**Table 1.** Genes associated with carbon metabolism in water hyacinth.

<table>
<thead>
<tr>
<th>Sequences producing significant alignments</th>
<th>Enzyme code</th>
<th>e-value</th>
<th>Hit length</th>
<th>Similarity mean%</th>
</tr>
</thead>
<tbody>
<tr>
<td>HK; hexokinase</td>
<td>2.7.1.1</td>
<td>3.01E-82</td>
<td>432</td>
<td>92.5</td>
</tr>
<tr>
<td>GCK; glucokinase</td>
<td>2.7.1.2</td>
<td>5.41E-63</td>
<td>420</td>
<td>87.3</td>
</tr>
<tr>
<td>ppgK; polyphosphate glucokinase</td>
<td>2.7.1.63</td>
<td>8.01E-02</td>
<td>419</td>
<td>78.4</td>
</tr>
<tr>
<td>ADPGK; ADP-dependent glucokinase</td>
<td>2.7.1.147</td>
<td>4.34E-123</td>
<td>428</td>
<td>93.4</td>
</tr>
<tr>
<td>pfkC; ADP-dependent phosphofructokinase/glucokinase</td>
<td>2.7.1.146</td>
<td>1.25E-138</td>
<td>407</td>
<td>84.9</td>
</tr>
<tr>
<td>pgi; glucose-6-phosphate isomerase</td>
<td>5.3.1.9</td>
<td>7.32E-12</td>
<td>441</td>
<td>60.4</td>
</tr>
<tr>
<td>pgil; glucose-6-phosphate isomerase, archaeal</td>
<td>5.3.1.9</td>
<td>6.57E-24</td>
<td>416</td>
<td>89.4</td>
</tr>
</tbody>
</table>
Table 2. Genes associated with fast growth

<table>
<thead>
<tr>
<th>Sequences producing significant alignments</th>
<th>Enzyme code</th>
<th>e-value</th>
<th>Hit length</th>
<th>Similarity mean%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondrial phosphate carrier mitochondrial like</td>
<td>1.5.3.8</td>
<td>6.09E-162</td>
<td>263</td>
<td>95.65</td>
</tr>
<tr>
<td>U1 small nuclear ribonucleoprotein 70 kDa</td>
<td>1.9.2.1</td>
<td>2.98E-86</td>
<td>453</td>
<td>87.45</td>
</tr>
<tr>
<td>RNA polymerase-associated LEO1</td>
<td>2.2.9.2</td>
<td>3.8E-84</td>
<td>386</td>
<td>84.2</td>
</tr>
<tr>
<td>Phosphoserine aminotransferase chloroplastic like</td>
<td>4.55E178</td>
<td>421</td>
<td></td>
<td>84.2</td>
</tr>
<tr>
<td>Inner membrane PPF-chloroplastic</td>
<td>1.5.4.8</td>
<td>5.51E-171</td>
<td>176</td>
<td>66.25</td>
</tr>
<tr>
<td>60s ribosomal L5</td>
<td>1.3.6.0</td>
<td>3.01E-82</td>
<td>256</td>
<td>68.35</td>
</tr>
<tr>
<td>Histone deacetylase HDT2-like</td>
<td>1.2.8.7</td>
<td>3E-72</td>
<td>457</td>
<td>66.25</td>
</tr>
<tr>
<td>Double-stranded RNA-binding 6</td>
<td>1.7.6.4</td>
<td>1.25E-138</td>
<td>423</td>
<td>89.2</td>
</tr>
<tr>
<td>26s proteasome non-ATpase regulatory subunit12 homolog A like</td>
<td>1.5.0.6</td>
<td>4.3E-34</td>
<td>471</td>
<td>96.0</td>
</tr>
<tr>
<td>Cytosolic NADP-malik enzyme</td>
<td>1.1.1.40</td>
<td>2.15E-32</td>
<td>418</td>
<td>85.4</td>
</tr>
</tbody>
</table>
The presence of similar gene sequences such as hexokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase and glucose/mannose-6-phosphate isomerase among others is proof that the water hyacinth is efficient in carbon metabolism as it can utilize any carbon source available.

Gene sequences shown in the Table 2 above like NADP-malic enzyme works in different pathways in plants. It can be involved in plant defense mechanisms for example as wound and UV-B radiation. The expression of the gene coding for cytosolic NADP-malic enzyme, by GenBank Accession Number AY444338 is also seen in rice (*Oryza sativa*) as well as in transgenic Arabidopsis. Rice a plant has remarkable success in water survival. The presence of Mitochondrial phosphate carrier mitochondrial like, U1 small nuclear ribonucleo 70 kDa, Inner membrane PPF-chloroplastic and other cellular component related genes confer to the fast growth and reproduction in the water hyacinth. 26s proteasome non-ATPase regulatory subunit 12 homolog A like and Cystolic NADP-malik enzyme found in the water hyacinth transcriptome are gene sequences associated with formation of air-filled roots. This confers that the water hyacinth uses these and other metabolism related genes to make the roots bulbous with many air sacs.

**Table 3.** Genes associated with leaves spatial arrangement and flowering.

<table>
<thead>
<tr>
<th>Sequences alignments producing significant alignments</th>
<th>Enzyme code</th>
<th>e-value</th>
<th>Hit length</th>
<th>Similarity mean%</th>
</tr>
</thead>
<tbody>
<tr>
<td>FT-like</td>
<td>2.1.3.12</td>
<td>2.01E-2.3</td>
<td>4896</td>
<td>82</td>
</tr>
<tr>
<td>Alpha glucagon water chloroplastic isoform x1</td>
<td>1.1.2.52</td>
<td>3.23E-1.2</td>
<td>3248</td>
<td>82.3</td>
</tr>
<tr>
<td>Arginine decarboxylase -like</td>
<td>3.1.1.8</td>
<td>8.3E-2.6</td>
<td>3065</td>
<td>94.4</td>
</tr>
<tr>
<td>2-oxoglutarate mitochondrial-like</td>
<td>1.1.1.32</td>
<td>7.32E-0.1</td>
<td>2398</td>
<td>67</td>
</tr>
<tr>
<td>Decussate</td>
<td>3.1.1.23</td>
<td>2.6E-3.5</td>
<td>1267</td>
<td>93.7</td>
</tr>
</tbody>
</table>
Jacalin-related lectin gene (EcJRL-1) & 1.4.1.21 & 5.3E-2.7 & 456 & 90.2 \\
AGL6 (Agamous-like 6)-like MADS-box gene, HoAGL6 & 1.1.1.14 & 3.4E-3.2 & 234 & 86.4 \\
ThiDE hydroxymethylpyrimidine kinase & 2.1.1.56 & 4.9E-8.9 & 437 & 75.3 \\

Ft-like and FTF family proteins coded for by the FT sequence shown in the Table above are predominantly involved in flowering of Arabidopsis thaliana (Conesa et al., 2005). Flowering is an important process in plants as it controls reproduction by timing correctly when plants should flower. Water hyacinths also have such a gene in their transcriptome which confers to their controlled and active flowering throughout their reproductive life cycle. The presence decussate gene in the water hyacinth controls spatial arrangement of leaves on a stem and enhances cell division as cited in the study of rice plant genes. It controls the phylotactic pattern by affecting cytokinins (McGinnis and Madden, 2004). Gene controlling the expression of hydroxymethylpyrimidine kinase which is responsible for metabolism of cofactors and vitamins.

**Table 4.** Genes associated with waxy cuticles formation.

<table>
<thead>
<tr>
<th>Sequences alignments</th>
<th>producing significant alignments</th>
<th>Enzyme code</th>
<th>e-value</th>
<th>Hit length</th>
<th>Similarity mean%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetyl-CoA acetyltransferase</td>
<td>2.1.1.52</td>
<td>2.03E-15</td>
<td>2038</td>
<td>72.7</td>
<td></td>
</tr>
<tr>
<td>Omega-3 fatty acid chloroplastic-like</td>
<td>1.1.2.32</td>
<td>2005</td>
<td>82.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clathrin assembly At5g35200</td>
<td>3.1.1.15</td>
<td>1869</td>
<td>80.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-oxoglutarate mitochondrial like</td>
<td>2.1.2.43</td>
<td>685</td>
<td>89.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UDP-N-acetylglucosamine-peptide acetylglucosamine</td>
<td>N-1.1.1.16</td>
<td>1701</td>
<td>93.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polyadenylate-binding RBP47-like</td>
<td>2.1.1.27</td>
<td>2.16E-23</td>
<td>1661</td>
<td>85.1</td>
<td></td>
</tr>
</tbody>
</table>
Biosynthesis of secondary metabolites such as fatty acid metabolism and pyruvate metabolism is controlled by the gene acetyl-CoA acetyltransferase. Homeobox leucine zipper controls several pathways which are aligned to the formation of cuticles is also found in the hits after a blast of the water hyacinth transcriptome. The genes listed confer to one of the features of water survival that is of waxy cuticles as they are synthesised by fatty acid metabolic pathways. Glucomannan 4-beta mannosyltransferase controls among other pathways the one which forms aerenchyma cells in roots and leaves, to help the water hyacinth float. This gene control has also been elucidated in other research with Arabidospsis (Conesa et al., 2005).

4.8. Summary.

The chapter highlighted the main results of blast, mapping and annotation. The results were presented mainly in pie charts and bar graphs which was followed by a brief description as well as the discussion. The next chapter, chapter 5 summarises the procedure, draws conclusions and recommendations basing on the results presented in chapter 4.
CHAPTER 5

5.0 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1. Introduction

In this chapter the researcher summarised the purpose of the study, and the main research findings. The research focused on establishing a list of genes and pathways responsible for the water hyacinth features that make it better adaptable to water as well as making it invasive. The conclusions drawn by the researcher were based on the results found in this study. Recommendations for further research and or repeating the study were also outlined.

5.2. Methods and Procedures

The researcher used an exploratory method to look up the online databases for similarities with the water hyacinth transcriptome. Blast2Go was used as the computer application software to manage the local alignment search task. First step was to search the online non-redundant datasets maintained by NCBI, a process known as “Basting”. Then comes mapping which looks up the specific gene positions and lastly annotation which leads to Analysis. The Blast program is directly linked to GO databases hence the name Blast2Go, which implies that after blast runs, the annotations are automatically linked to GO.

5.3. Major Findings

The main findings of this research are that the genes responsible for the making the water hyacinth invasive with respect to buoyancy, waxy cuticles on leaves, fast growth rate and efficient reproduction were listed. Other findings are that those genes in the water hyacinth which had hits were furnished with annotation information to show the related biological pathways which they
encode. Classification of biological pathways into biological processes, molecular function and cellular component was also done. A comparison of the water hyacinth transcriptome with the genes and pathways from other plants such as *Arabidopsis thaliana* and *Oryza sativa* was done.

5.4. **Conclusion**

From the results obtained and presented in chapter 4, it can be concluded that the list of genes associated with several biological pathways is key to inform biotechnology laboratory work to either switch off the genes as it were or to inform design of herbicides to target those specific genes, pathways and or enzymes involved thereof.

5.5. **Recommendations**

Since this is a novel area of study, the researcher after citing some challenges, highlighted in chapter 3, strongly recommend that this research be repeated by a group of students to share the cumbersome work. It is also recommended that this research be done in parallel with real time PCR for verification of the research results, since this was entirely a desktop and online project.

It is further recommended that extraction of the genes be done by laboratory work on sequencing to augment the major findings, since the actual practical sequencing could not be done as the researcher did not have enough funding. Further research is also required to investigate those sequences which did not have blast hits, not found in KEGG analysis and those which did not have any annotation information, or to be curated.
6.0 REFERENCES.


Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics 

1016/j.chemosphere.2006.06.04109.020.

12. FAO (Food and Agriculture Organization of the United Nations). (2013). Management of 

Water hyacinth (Eichhornia crassipes) dynamics and Succession in the Nyanza Gulf of Lake 
Victoria (East Africa): Implications for Water Quality and Biodiversity, Conservation, 
Scientific World Journal.

Congreso Nacional sobre Especies Exoticas Invasoras, León 19-22Septiembre: 118.

Sciences, Roskild Universitetscenter Nib 14.2.

insects and plant pathogens in Mexico. Crop Protection 26: 1234-1238, 
http://dx.doi.org/10.1016/j.cropro.2006.10.028.


APPENDIX

Appendix 1. KO terms list.

ko:K00085 ko:K00095 ko:K00142 ko:K00311 ko:K00327 ko:K00555 ko:K00573 ko:K00670
ko:K06072 ko:K06158 ko:K06174 ko:K06184 ko:K06185 ko:K06890 ko:K06891 ko:K06911
ko:K09338 ko:K09419 ko:K09422 ko:K09493 ko:K09494 ko:K09495 ko:K09496 ko:K09497
ko:K09660 ko:K09716 ko:K09872 ko:K09873 ko:K09874 ko:K09875 ko:K10257 ko:K10268
Appendix 2. KEGG mapper results.
Appendix 4. The water hyacinth picture.
Appendix 5. Blast2Go logo.