ABOUT AJB

The **African Journal of Biotechnology (AJB)** (ISSN 1684-5315) is published weekly (one volume per year) by Academic Journals.

African Journal of Biotechnology (AJB), a new broad-based journal, is an open access journal that was founded on two key tenets: To publish the most exciting research in all areas of applied biochemistry, industrial microbiology, molecular biology, genomics and proteomics, food and agricultural technologies, and metabolic engineering. Secondly, to provide the most rapid turn-around time possible for reviewing and publishing, and to disseminate the articles freely for teaching and reference purposes. All articles published in AJB are peer-reviewed.

Submission of Manuscript

Please read the **Instructions for Authors** before submitting your manuscript. The manuscript files should be given the last name of the first author

Click here to Submit manuscripts online

If you have any difficulty using the online submission system, kindly submit via this email ajb@academicjournals.org.

With questions or concerns, please contact the Editorial Office at ajb@academicjournals.org.

Dr. Alfred Dixon
International Institute of Tropical Agriculture (IITA)
PMB 5320, Ibadan
Oyo State, Nigeria

Dr. Sankale Shompole
Dept. of Microbiology, Molecular Biology and Biochemisty,
University of Idaho, Moscow,
ID 83844, USA.

Dr. Mathew M. Abang
Germplasm Program
International Center for Agricultural Research in the Dry Areas
(ICARDA)
P.O. Box 5466, Aleppo, SYRIA.

Dr. Solomon Olawale Odemuyiwa
Pulmonary Research Group
Department of Medicine
550 Heritage Medical Research Centre
University of Alberta
Edmonton
Canada T6G 2S2

Prof. Anna-Maria Botha-Oberholster
Plant Molecular Genetics
Department of Genetics
Forestry and Agricultural Biotechnology Institute
Faculty of Agricultural and Natural Sciences
University of Pretoria
ZA-0002 Pretoria, South Africa

Dr. O. U. Ezeronye
Department of Biological Science
Michael Okpara University of Agriculture
Umudike, Abia State, Nigeria.

Dr. Joseph Hounhouigan
Maître de Conférence
Sciences et technologies des aliments
Faculté des Sciences Agronomiques
Université d'Abomey-Calavi
01 BP 526 Cotonou
République du Bénin

Prof. Christine Rey
Dept. of Molecular and Cell Biology,
University of the Witwatersand,
Private Bag 3, WITS 2050, Johannesburg, South Africa

Dr. Kamel Ahmed Abd-Elsalam
Molecular Markers Lab. (MML)
Plant Pathology Research Institute (PPathRI)
Agricultural Research Center, 9-Gamma St., Orman, 12619,
Giza, Egypt

Dr. Jones Lemchi
International Institute of Tropical Agriculture (IITA)
Onne, Nigeria

Prof. Greg Blatch
Head of Biochemistry & Senior Wellcome Trust Fellow
Department of Biochemistry, Microbiology & Biotechnology
Rhodes University
Grahamstown 6140
South Africa

Dr. Beatrice Kilel
P.O Box 1413
Manassas, VA 20108
USA

Dr. Jackie Hughes
Research-for-Development
International Institute of Tropical Agriculture (IITA)
Ibadan, Nigeria

Dr. Robert L. Brown
Southern Regional Research Center,
U.S. Department of Agriculture,
Agricultural Research Service,
New Orleans, LA 70179.

Dr. Deborah Rayfield
Physiology and Anatomy
Bowie State University
Department of Natural Sciences
Crawford Building, Room 003C
Bowie MD 20715,USA

Dr. Marlene Shehata
University of Ottawa Heart Institute
Genetics of Cardiovascular Diseases
40 Ruskin Street
K1Y-4W7, Ottawa, ON, CANADA

Dr. Hany Sayed Hafez
The American University in Cairo,
Egypt

Dr. Clement O. Adebooye
Department of Plant Science
Obafemi Awolowo University, Ile-Ife
Nigeria

Dr. Ali Demir Sezer
Marmara Üniversitesi Eczacilik Fakültesi,
Tibbiye cad. No: 49, 34668, Haydarpasa, Istanbul,
Turkey

Dr. Ali Gazanchain
P.O. Box: 91735-1148, Mashhad,
Iran.

Dr. Anant B. Patel
Centre for Cellular and Molecular Biology
Uppal Road, Hyderabad 500007
India

Prof. Arne Elofsson
Department of Biophysics and Biochemistry
Bioinformatics at Stockholm University,
Sweden

Prof. Bahram Goliaei
Departments of Biophysics and Bioinformatics
Laboratory of Biophysics and Molecular Biology
University of Tehran, Institute of Biochemistry
and Biophysics
Iran

Dr. Nora Babudri
Dipartimento di Biologia cellulare e ambientale
Università di Perugia
Via Pascoli
Italy

Dr. S. Adesola Ajayi
Seed Science Laboratory
Department of Plant Science
Faculty of Agriculture
Obafemi Awolowo University
Ile-Ife 220005, Nigeria

Dr. Yee-Joo TAN
Department of Microbiology
Yong Loo Lin School of Medicine,
National University Health System (NUHS),
National University of Singapore
MD4, 5 Science Drive 2,
Singapore 117597
Singapore

Prof. Hidetaka Hori
Laboratories of Food and Life Science,
Graduate School of Science and Technology,
Niigata University.
Niigata 950-2181,
Japan

Prof. Thomas R. DeGregori
University of Houston,
Texas 77204 5019,
USA

Dr. Wolfgang Ernst Bernhard Jelkmann
Medical Faculty, University of Lübeck,
Germany

Dr. Moktar Hamdi
Department of Biochemical Engineering,
Laboratory of Ecology and Microbial Technology
National Institute of Applied Sciences and
Technology.
BP: 676. 1080,
Tunisia

Dr. Salvador Ventura
Department de Bioquímica i Biologia Molecular
Institut de Biotecnologia i de Biomedicina
Universitat Autònoma de Barcelona
Bellaterra-08193
Spain

Dr. Claudio A. Hetz
Faculty of Medicine, University of Chile
Independencia 1027
Santiago, Chile

Prof. Felix Dapare Dakora
Research Development and Technology Promotion
Cape Peninsula University of Technology,
Room 2.8 Admin. Bldg. Keizersgracht, P.O. 652,
Cape Town 8000,
South Africa

Dr. Geremew Bultosa
Department of Food Science and Post harvest Technology
Haramaya University
Personal Box 22, Haramaya University Campus
Dire Dawa,
Ethiopia

Dr. José Eduardo Garcia
Londrina State University
Brazil

Prof. Nirbhay Kumar
Malaria Research Institute
Department of Molecular Microbiology and Immunology
Johns Hopkins Bloomberg School of Public Health
E5144, 615 N. Wolfe Street
Baltimore, MD 21205

Prof. M. A. Awal
Department of Anatomy and Histplogy,
Bangladesh Agricultural University,
Mymensingh-2202,
Bangladesh

Prof. Christian Zwieb
Department of Molecular Biology
University of Texas Health Science Center at Tyler
11937 US Highway 271
Tyler, Texas 75708-3154
USA

Prof. Danilo López-Hernández
Instituto de Zoología Tropical, Facultad de Ciencias,
Universidad Central de Venezuela.
Institute of Research for the Development (IRD),
Montpellier,
France

Prof. Donald Arthur Cowan
Department of Biotechnology,
University of the Western Cape Bellville 7535
Cape Town,
South Africa

Dr. Ekhaise Osaro Frederick
University Of Benin, Faculty of Life Science
Department of Microbiology
P. M. B. 1154, Benin City, Edo State,
Nigeria.

Dr. Luísa Maria de Sousa Mesquita Pereira
IPATIMUP R. Dr. Roberto Frias, s/n 4200-465 Porto
Portugal

Dr. Min Lin
Animal Diseases Research Institute
Canadian Food Inspection Agency
Ottawa, Ontario,
Canada K2H 8P9

Prof. Nobuyoshi Shimizu
Department of Molecular Biology,
Center for Genomic Medicine
Keio University School of Medicine,
35 Shinanomachi, Shinjuku-ku
Tokyo 160-8582,
Japan

Dr. Adewunmi Babatunde Idowu
Department of Biological Sciences
University of Agriculture Abia
Abia State,
Nigeria

Dr. Yifan Dai
Associate Director of Research
Revivicor Inc.
100 Technology Drive, Suite 414
Pittsburgh, PA 15219
USA

Dr. Zhongming Zhao
Department of Psychiatry, PO Box 980126,
Virginia Commonwealth University School of Medicine,
Richmond, VA 23298-0126,
USA

Prof. Giuseppe Novelli
Human Genetics,
Department of Biopathology,
Tor Vergata University, Rome,
Italy

Dr. Moji Mohammadi
402-28 Upper Canada Drive
Toronto, ON, M2P 1R9 (416) 512-7795
Canada

Prof. Jean-Marc Sabatier
Directeur de Recherche Laboratoire ERT-62
Ingénierie des Peptides à Visée Thérapeutique,
Université de la Méditerranée-Ambrilia
Biopharma inc.,
Faculté de Médecine Nord, Bd Pierre Dramard,
13916,
Marseille cédex 20.
France

Dr. Fabian Hoti
PneumoCarr Project
Department of Vaccines
National Public Health Institute
Finland

Prof. Irina-Draga Caruntu
Department of Histology
Gr. T. Popa University of Medicine and Pharmacy
16, Universitatii Street, Iasi,
Romania

Dr. Dieudonné Nwaga
Soil Microbiology Laboratory,
Biotechnology Center. PO Box 812,
Plant Biology Department,
University of Yaoundé I, Yaoundé,
Cameroon

Dr. Gerardo Armando Aguado-Santacruz
Biotechnology CINVESTAV-Unidad Irapuato
Departamento Biotecnología
Km 9.6 Libramiento norte Carretera Irapuato-
León Irapuato,
Guanajuato 36500
Mexico

Dr. Abdolkaim H. Chehregani
Department of Biology
Faculty of Science
Bu-Ali Sina University
Hamedan,
Iran

Dr. Abir Adel Saad
Molecular oncology
Department of Biotechnology
Institute of graduate Studies and Research
Alexandria University,
Egypt

Dr. Azizul Baten
Department of Statistics
Shah Jalal University of Science and Technology
Sylhet-3114,
Bangladesh

Dr. Bayden R. Wood
Australian Synchrotron Program
Research Fellow and Monash Synchrotron
Research Fellow Centre for Biospectroscopy
School of Chemistry Monash University Wellington
Rd. Clayton,
3800 Victoria,
Australia

Dr. G. Reza Balali
Molecular Mycology and Plant Pthology
Department of Biology
University of Isfahan
Isfahan
Iran

Dr. Beatrice Kilel
P.O Box 1413
Manassas, VA 20108
USA

Prof. H. Sunny Sun
Institute of Molecular Medicine
National Cheng Kung University Medical College
1 University road Tainan 70101,
Taiwan

Prof. Ima Nirwana Soelaiman
Department of Pharmacology
Faculty of Medicine
Universiti Kebangsaan Malaysia
Jalan Raja Muda Abdul Aziz
50300 Kuala Lumpur,
Malaysia

Prof. Tunde Ogunsanwo
Faculty of Science,
Olabisi Onabanjo University,
Ago-Iwoye.
Nigeria

Dr. Evans C. Egwim
Federal Polytechnic,
Bida Science Laboratory Technology Department,
PMB 55, Bida, Niger State,
Nigeria

Prof. George N. Goulielmos
Medical School,
University of Crete
Voutes, 715 00 Heraklion, Crete,
Greece

Dr. Uttam Krishna
Cadila Pharmaceuticals limited ,
India 1389, Tarsad Road,
Dholka, Dist: Ahmedabad, Gujarat,
India

Prof. Mohamed Attia El-Tayeb Ibrahim
Botany Department, Faculty of Science at Qena,
South Valley University, Qena 83523,
Egypt

Dr. Nelson K. Ojijo Olang'o
Department of Food Science & Technology,
JKUAT P. O. Box 62000, 00200, Nairobi,
Kenya

Dr. Pablo Marco Veras Peixoto
University of New York NYU College of Dentistry
345 E. 24th Street, New York, NY 10010
USA

Prof. T E Cloete
University of Pretoria Department of
Microbiology and Plant Pathology,
University of Pretoria,
Pretoria,
South Africa

Prof. Djamel Saidi
Laboratoire de Physiologie de la Nutrition et de
Sécurité
Alimentaire Département de Biologie,
Faculté des Sciences,
Université d'Oran, 31000 - Algérie
Algeria

Dr. Tomohide Uno
Department of Biofunctional chemistry,
Faculty of Agriculture Nada-ku,
Kobe., Hyogo, 657-8501,
Japan

Dr. Ulises Urzúa
Faculty of Medicine,
University of Chile Independencia 1027, Santiago,
Chile

Dr. Aritua Valentine
National Agricultural Biotechnology Center,
Kawanda
Agricultural Research Institute (KARI)
P.O. Box, 7065, Kampala,
Uganda

Prof. Yee-Joo Tan
Institute of Molecular and Cell Biology 61 Biopolis
Drive,
Proteos, Singapore 138673
Singapore

Prof. Viroj Wiwanitkit
Department of Laboratory Medicine,
Faculty of Medicine, Chulalongkorn University,
Bangkok
Thailand

Dr. Thomas Silou
Universit of Brazzaville BP 389
Congo

Prof. Burtram Clinton Fielding
University of the Western Cape
Western Cape,
South Africa

Dr. Brnčić (Brncic) Mladen
Faculty of Food Technology and Biotechnology,
Pierottijeva 6,
10000 Zagreb,
Croatia.

Dr. Meltem Sesli
College of Tobacco Expertise,
Turkish Republic, Celal Bayar University 45210,
Akhisar, Manisa,
Turkey.

Dr. Idress Hamad Attitalla
Omar El-Mukhtar University,
Faculty of Science,
Botany Department,
El-Beida, Libya.

Dr. Linga R. Gutha
Washington State University at Prosser,
24106 N Bunn Road,
Prosser WA 99350-8694.

Dr Helal Ragab Moussa
Bahnay, Al-bagour, Menoufia,
Egypt.

Dr VIPUL GOHEL
DuPont Industrial Biosciences
Danisco (India) Pvt Ltd
5th Floor, Block 4B,
DLF Corporate Park
DLF Phase III
Gurgaon 122 002
Haryana (INDIA)

Dr. Sang-Han Lee
Department of Food Science & Biotechnology,
Kyungpook National University
Daegu 702-701,
Korea.

Dr. Bhaskar Dutta
DoD Biotechnology High Performance Computing
Software Applications
Institute (BHSAI)
U.S. Army Medical Research and Materiel
Command
2405 Whittier Drive
Frederick, MD 21702

Dr. Muhammad Akram
Faculty of Eastern Medicine and Surgery,
Hamdard Al-Majeed College of Eastern Medicine,
Hamdard University,
Karachi.

Dr. M. Muruganandam
Departtment of Biotechnology
St. Michael College of Engineering & Technology,
Kalayarkoil,
India.

Dr. Gökhan Aydin
Suleyman Demirel University,
Atabey Vocational School,
Isparta-Türkiye,

Dr. Rajib Roychowdhury
Centre for Biotechnology (CBT),
Visva Bharati,
West-Bengal,
India.

Dr Takuji Ohyama
Faculty of Agriculture, Niigata University

Dr Mehdi Vasfi Marandi
University of Tehran

Dr FÜgen DURLU-ÖZKAYA
Gazi Üniversity, Tourism Faculty, Dept. of
Gastronomy and Culinary Art

Dr. Reza Yari
Islamic Azad University, Boroujerd Branch

Dr Zahra Tahmasebi Fard
Roudehen branche, Islamic Azad University

Dr Albert Magrí
Giro Technological Centre

Dr Ping ZHENG
Zhejiang University, Hangzhou, China

Dr. Kgomotso P. Sibeko
University of Pretoria

Dr Greg Spear
Rush University Medical Center

Prof. Pilar Morata
University of Malaga

Dr Jian Wu
Harbin medical university , China

Dr Hsiu-Chi Cheng
National Cheng Kung University and Hospital.

Prof. Pavel Kalac
University of South Bohemia, Czech Republic

Dr Kürsat Korkmaz
Ordu University, Faculty of Agriculture,
Department of Soil Science and Plant Nutrition

Dr. Shuyang Yu
Department of Microbiology, University of Iowa
Address: 51 newton road, 3-730B BSB bldg. Iowa
City, IA, 52246, USA

Dr. Binxing Li

Dr. Mousavi Khaneghah
College of Applied Science and Technology-
Applied Food Science, Tehran, Iran.

Dr. Qing Zhou
Department of Biochemistry and Molecular
Biology,
Oregon Health and Sciences University Portland.

Dr Legesse Adane Bahiru
Department of Chemistry,
Jimma University,
Ethiopia.

Dr James John
School Of Life Sciences,
Pondicherry University,
Kalapet, Pondicherry

Instructions for Author

Electronic submission of manuscripts is strongly encouraged, provided that the text, tables, and figures are included in a single Microsoft Word file (preferably in Arial font).

The **cover letter** should include the corresponding author's full address and telephone/fax numbers and should be in an e-mail message sent to the Editor, with the file, whose name should begin with the first author's surname, as an attachment.

Article Types
Three types of manuscripts may be submitted:

Regular articles: These should describe new and carefully confirmed findings, and experimental procedures should be given in sufficient detail for others to verify the work. The length of a full paper should be the minimum required to describe and interpret the work clearly.

Short Communications: A Short Communication is suitable for recording the results of complete small investigations or giving details of new models or hypotheses, innovative methods, techniques or apparatus. The style of main sections need not conform to that of full-length papers. Short communications are 2 to 4 printed pages (about 6 to 12 manuscript pages) in length.

Reviews: Submissions of reviews and perspectives covering topics of current interest are welcome and encouraged. Reviews should be concise and no longer than 4-6 printed pages (about 12 to 18 manuscript pages). Reviews are also peer-reviewed.

Review Process

All manuscripts are reviewed by an editor and members of the Editorial Board or qualified outside reviewers. Authors cannot nominate reviewers. Only reviewers randomly selected from our database with specialization in the subject area will be contacted to evaluate the manuscripts. The process will be blind review.
Decisions will be made as rapidly as possible, and the journal strives to return reviewers' comments to authors as fast as possible. The editorial board will re-review manuscripts that are accepted pending revision. It is the goal of the AJFS to publish manuscripts within weeks after submission.

Regular articles

All portions of the manuscript must be typed double-spaced and all pages numbered starting from the title page.

The Title should be a brief phrase describing the contents of the paper. The Title Page should include the authors' full names and affiliations, the name of the corresponding author along with phone, fax and E-mail information. Present addresses of authors should appear as a footnote.

The Abstract should be informative and completely self-explanatory, briefly present the topic, state the scope of the experiments, indicate significant data, and point out major findings and conclusions. The Abstract should be 100 to 200 words in length.. Complete sentences, active verbs, and the third person should be used, and the abstract should be written in the past tense. Standard nomenclature should be used and abbreviations should be avoided. No literature should be cited.
Following the abstract, about 3 to 10 key words that will provide indexing references should be listed.

A list of non-standard **Abbreviations** should be added. In general, non-standard abbreviations should be used only when the full term is very long and used often. Each abbreviation should be spelled out and introduced in parentheses the first time it is used in the text. Only recommended SI units should be used. Authors should use the solidus presentation (mg/ml). Standard abbreviations (such as ATP and DNA) need not be defined.

The Introduction should provide a clear statement of the problem, the relevant literature on the subject, and the proposed approach or solution. It should be understandable to colleagues from a broad range of scientific disciplines.

Materials and methods should be complete enough to allow experiments to be reproduced. However, only truly new procedures should be described in detail; previously published procedures should be cited, and important modifications of published procedures should be mentioned briefly. Capitalize trade names and include the manufacturer's name and address. Subheadings should be used. Methods in general use need not be described in detail.

Results should be presented with clarity and precision. The results should be written in the past tense when describing findings in the authors' experiments. Previously published findings should be written in the present tense. Results should be explained, but largely without referring to the literature. Discussion, speculation and detailed interpretation of data should not be included in the Results but should be put into the Discussion section.

The Discussion should interpret the findings in view of the results obtained in this and in past studies on this topic. State the conclusions in a few sentences at the end of the paper. The Results and Discussion sections can include subheadings, and when appropriate, both sections can be combined.

The Acknowledgments of people, grants, funds, etc should be brief.

Tables should be kept to a minimum and be designed to be as simple as possible. Tables are to be typed double-spaced throughout, including headings and footnotes. Each table should be on a separate page, numbered consecutively in Arabic numerals and supplied with a heading and a legend. Tables should be self-explanatory without reference to the text. The details of the methods used in the experiments should preferably be described in the legend instead of in the text. The same data should not be presented in both table and graph form or repeated in the text.

Figure legends should be typed in numerical order on a separate sheet. Graphics should be prepared using applications capable of generating high resolution GIF, TIFF, JPEG or Powerpoint before pasting in the Microsoft Word manuscript file. Tables should be prepared in Microsoft Word. Use Arabic numerals to designate figures and upper case letters for their parts (Figure 1). Begin each legend with a title and include sufficient description so that the figure is understandable without reading the text of the manuscript. Information given in legends should not be repeated in the text.

References: In the text, a reference identified by means of an author's name should be followed by the date of the reference in parentheses. When there are more than two authors, only the first author's name should be mentioned, followed by 'et al'. In the event that an author cited has had two or more works published during the same year, the reference, both in the text and in the reference list, should be identified by a lower case letter like 'a' and 'b' after the date to distinguish the works.

Examples:

Abayomi (2000), Agindotan et al. (2003), (Kelebeni, 1983), (Usman and Smith, 1992), (Chege, 1998; 1987a,b; Tijani, 1993,1995), (Kumasi et al., 2001)

References should be listed at the end of the paper in alphabetical order. Articles in preparation or articles submitted for publication, unpublished observations, personal communications, etc. should not be included in the reference list but should only be mentioned in the article text (e.g., A. Kingori, University of Nairobi, Kenya, personal communication). Journal names are abbreviated according to Chemical Abstracts. Authors are fully responsible for the accuracy of the references.

Examples:

Chikere CB, Omoni VT and Chikere BO (2008). Distribution of potential nosocomial pathogens in a hospital environment. Afr. J. Biotechnol. 7: 3535-3539.

Moran GJ, Amii RN, Abrahamian FM, Talan DA (2005). Methicillinresistant Staphylococcus aureus in community-acquired skin infections. Emerg. Infect. Dis. 11: 928-930.

Pitout JDD, Church DL, Gregson DB, Chow BL, McCracken M, Mulvey M, Laupland KB (2007). Molecular epidemiology of CTXM-producing Escherichia coli in the Calgary Health Region: emergence of CTX-M-15-producing isolates. Antimicrob. Agents Chemother. 51: 1281-1286.

Pelczar JR, Harley JP, Klein DA (1993). Microbiology: Concepts and Applications. McGraw-Hill Inc., New York, pp. 591-603.

Short Communications

Short Communications are limited to a maximum of two figures and one table. They should present a complete study that is more limited in scope than is found in full-length papers. The items of manuscript preparation listed above apply to Short Communications with the following differences: (1) Abstracts are limited to 100 words; (2) instead of a separate Materials and Methods section, experimental procedures may be incorporated into Figure Legends and Table footnotes; (3) Results and Discussion should be combined into a single section.

Proofs and Reprints: Electronic proofs will be sent (e-mail attachment) to the corresponding author as a PDF file. Page proofs are considered to be the final version of the manuscript. With the exception of typographical or minor clerical errors, no changes will be made in the manuscript at the proof stage.

African Journal of Biotechnology

Table of Contents: Volume 13 Number 21, 21 May, 2014

ARTICLES

Table of Contents: Volume 13 Number 21, 21 May, 2014

Full Length Research Paper

Dearth of full-length HIV-1 sequences obscures the true HIV-1 genetic subtypes distribution in sub-Saharan Africa

Kemebradikumo Pondei*, Abdulrasheed Abdu and Langley Orutugu

Department of Medical Microbiology and Parasitology, Faculty of Basic Medical Sciences, College of Health Sciences, Niger Delta University, Amassoma, Wilberforce Island, Bayelsa State, Nigeria.

HIV infection is still a public health problem in sub-Saharan Africa. The broad diversity exhibited by HIV-1 may impact on transmission, disease progression, drug resistance and vaccine development. Most analyses of HIV-1 subtype distribution have been on partial *HIV-1* gene sequences, which may not adequately reflect the circulating subtypes. The objective of this study was to estimate the HIV-1 subtype distribution in sub-Saharan Africa using only full-length genome sequences. Using available HIV-1 full-length genome sequences from sub-Saharan Africa, the HIV-1 distribution in the region was analysed and compared with a previous global analysis which was not based entirely on full-length sequences. A total of 934 HIV-1 full-length genome sequences were available from 27 sub-Saharan countries. There was a disproportionate distribution of HIV-1 subtypes among countries with Cameroon having all the four HIV-1 groups. The subtype C was the most available in addition to a large proportion of circulating and unique recombinant forms (CRFs/URFs) especially in Central and West African countries, with frequencies of 32.6 to 90%. There was decreased representation of subtypes A and G in regions where CRFs/URFs were common compared with previous analysis using partial sequences. There is a need for more HIV-1 full-length genome sequences from sub-Saharan Africa for the true distribution of HIV-1 subtypes to be known, as analysis of partial sequences is not truly representative of the circulating subtypes.

Key words: Africa, distribution, genetic diversity, HIV sequence variability, subtypes, recombination.

INTRODUCTION

Infection with the Human Immunodeficiency Virus (HIV) continues to be a global public health problem with devastating consequences in developing countries especially in sub-Saharan Africa, even though antiretroviral therapy has improved the quality of life of those infected. HIV exists in two genetically distinct forms (HIV-1 and HIV-2), with HIV-2 being restricted to West Africa and HIV-1 having a global spread and being responsible for the HIV pandemic.

HIV-1 exhibits genetic diversity in the form of viral

quasispecies (Meyerhans et al., 1989) described as a heterogeneous viral population of related genomes (Domingo et al., 1997). This genetic diversity of HIV-1 is believed to result from a high mutation rate due to the infidelity or error-prone characteristic of reverse transcriptase during replication (Roberts et al., 1988; Boyer et al., 1992), a high replication rate of about 10^9 virions per day (Ho et al., 1995) and genomic recombination (Hu and Temin, 1990; Jetzt et al., 2000; Zhuang et al., 2002).

HIV-1 is classified into three main genetic groupings each representing independent cross-species trans-mission, although a fourth group (Group P) (Plantier et al., 2009) has been suggested. The three groups are: Group M (major), Group O (outlier) and Group N (non-M, non-O). Groups O and N are mainly restricted to Cameroon and the Democratic Republic of Congo. The Group M has a global distribution and is further divided into nine subtypes and some sub-subtypes: subtypes A, B, C, D, F, G, H, J and K, with sub-subtypes A1 to A5 (Gao et al., 2001; Meloni et al., 2004; Vidal et al., 2006, 2009), and F1 and F2 (Triques et al., 1999). Combinations of two or more subtypes and/or sub-subtypes exist, and when these mosaic forms become widely spread and fixed in the population, they are known as circulating recombinant forms (CRFs). A CRF is therefore defined as "intersubtype recombination for which at least three epidemiologically unlinked variants are monophyletic and share identical genetic structure along their full genomes" (Yebra et al., 2012), while a unique recombinant form (URF) is a variant that has not been isolated from three or more individuals.

Presently, about 58 CRFs have been characterized and are in the public domain (http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CR Fs.html). HIV-1 subtypes may exhibit phenotypic differences. Subtypes are believed to impact on tropism, with some studies associating increased CXCR4 usage with infections with subtype C (Johnston et al., 2003; Connell et al., 2008), and others finding decreased CXCR4 usage in subtype C infections (Bjorndal et al., 1997; Abebe et al., 1999; Peeters et al., 1999; Esbjornsson et al., 2010). Subtypes may have an important effect on transmission of HIV-1, as the subtype B was associated with homosexual transmission and the subtype C with heterosexual transmission (van Harmelen et al., 1997; van Harmelen et al., 2001), but a heterosexually driven subtype B epidemic has been observed in Trinidad and Tobago (Cleghorn et al., 2000). Also, the subtype C is believed to be more likely to be transmitted vertically than the subtypes A and D (Blackard et al., 2001; Renjifo et al., 2001; Renjifo et al., 2003); and infection with the subtype D has been associated with faster CD4 T cell decline and a faster rate of disease progression (Alaeus et al., 1999; Kanki et al., 1999; Kaleebu et al., 2001; Vasan et al., 2006; Baeten et al., 2007; Kaleebu et al., 2007; Easterbrook et

al., 2010). Subtypes could also be important in vaccine design and development (Hemelaar et al., 2011).

Comprehensive understanding of disease process and effective interventions could arise from correlating genomic profile of HIV with that of patients (Sampathkumar et al., 2012). With HIV subtypes playing important roles in transmission and outcome of disease, it becomes imperative for the actual distribution of subtypes to be known. HIV subtype distribution has largely been determined using partial genome sequences. Since recombinant forms are relatively common, there is the possibility that the actual subtype distribution is not accurately represented using these partial genome sequences, and recombinants could be artificially scored as pure subtypes. This analysis therefore sought to firstly determine the HIV-1 subtype distribution in sub-Saharan Africa using only full-length or near full-length HIV-1 genome sequences available in the public domain. Secondly, it intended to document available full-length HIV-1 sequences from the region.

MATERIALS AND METHODS

All available full-length or near full-length HIV-1 genome sequences of sub-Saharan African origin were obtained from GenBank (http://www.ncbi.nlm.nih.gov) and the Los Alamos HIV Sequence Database (http://www.hiv.lanl.gov). The full-length sequence is one that contains the entire protein coding region as well as the non-coding regions, while a near full-length sequence contains almost all of the coding region.

Duplications were rectified with only one sequence per patient, except for those with infection with viruses of different subtypes. Information was transferred to an Excel spreadsheet and analysed. The countries in sub-Saharan Africa were grouped into four regions using the United Nations geoscheme for Africa: Central Africa, East Africa, Southern Africa and West Africa.

For convenience, Malawi, Zambia and Zimbabwe were grouped under Southern Africa.

Central Africa includes Angola, Cameroon, Central African Republic (CAR), Chad, Democratic Republic of Congo (DRC), Equatorial Guinea, Gabon, Republic of Congo and Sao Tome and Principe.

East Africa includes Burundi, Djibouti, Ethiopia, Eritrea, Kenya, Madagascar, Mauritius, Mozambique, Rwanda, Somalia, Sudan, Tanzania and Uganda.

Southern Africa includes Botswana, Lesotho, Malawi, Namibia, South Africa, Swaziland, Zambia and Zimbabwe.

West Africa includes Benin, Burkina Faso, Cape Verde, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone and Togo.

Results from this analysis were then compared with those obtained in an earlier analysis which did not discriminate between partial and full-length HIV genome sequences (Hemelaar et al., 2006).

RESULTS

A total of 1084 full-length or near full-length HIV-1 genome sequences of sub-Saharan Africa origin, were retrieved from GenBank (http://www.ncbi.nlm.nih.gov)

and the Los Alamos HIV Sequence Database (http://www.hiv.lanl.gov). The sequences had been submitted to the databases between 1983 and 2011. In contrast, there were 127,798 partial HIV-1 sequences from the region.

With the removal of duplications, there were 934 unique full-length HIV-1 sequences from patients within the sub-Saharan Africa region. Sequences were available from 27 of the 45 sub-Saharan countries. An overwhelming majority of the sequences (96.7%) were of the HIV-1 Group M, with Groups N and O accounting for 1.1 and 2.1%, respectively of the sequences (Table 1). The subtype C was the most common subtype accounting for almost half of the Group M sequences, with recombinants responsible for 29% of the Group M sequences. A global HIV-1 subtype analysis was done in 2004 (Hemelaar et al., 2006) and some of the results are compared with the present analysis (Table 1).

Regional analysis

Central Africa

There were available sequences from five of nine countries. Over three-quarters of the sequences were from Cameroon. Recombinant forms made up 64.2% of the sequences (Table 2). All the HIV groups (M, N, O and P) were present among the Cameroonian sequences, and also six of the nine Group M subtypes. Recombinants made up 66.7% of the sequences from Cameroon. The Group O was present among sequences from Gabon, while recombinants accounted for 66.7% of the sequences from the Democratic Republic of Congo.

East Africa

Sequences were retrieved for seven out of 13 countries in the region. Most of the sequences (94.4%) were from three countries: Kenya, Uganda and Tanzania. The subtype distribution of sequences from East Africa was spread between recombinants (37.2%), subtype A (32%), subtype D (18.4%) and subtype C (11.2%). About half of the sequences from Kenya were of the subtype A, with recombinants making up 41.8%. Of the Ugandan sequences, 44.2 and 32.6% were subtype D and recombinants respectively, while subtype A accounted for 23.2%. Of the sequences from Tanzania, 44% each were recombinants and the subtype C.

Southern Africa

Sequences were available for four out of eight countries in the region. Most of the sequences were from South Africa (78.4%) and 97.6% of the sequences from South

Africa were of the subtype C, with recombinants contributing just 1.2% of the sequences. All the sequences from Malawi and 94.7% of those from Zambia were also of the subtype C.

West Africa

Sequences were obtained for 10 out of 16 countries in the region. Over 75% of the sequences were from just three countries: Nigeria, Ghana and Senegal. In this region, 71.83% of the sequences were recombinants. The HIV-1 Group O was surprisingly present among the sequences from Senegal (Table 2). Among the Nigerian sequences, subtype G and recombinants were each responsible for 47.6% of the sequences. Recombinants were responsible for 90% of the sequences from Ghana and 46.1% of those from Senegal.

Circulating and unique recombinant forms (CRFs/URFs)

Recombinants accounted for a substantial proportion of the available sequences from certain countries in the West, Central and Eastern African regions (Table 3) ranging from 32.6 to 90%. Five countries had no recombinants (Chad, Djibouti, Ethiopia, Somalia and Malawi). Twenty-one out of the characterized 58 CRFs were present among the sequences. In terms of proportion, Ghana had the highest proportion of recombinant sequences, but Cameroon numerically had more CRFs/URFs.

Cameroon also had the greatest diversity of recombinants with nine different CRFs and 23 different URF types, followed by Kenya with 11 URF types and 2 CRFs; Democratic Republic of Congo (4 CRFs, 5 URFs) and Ghana (3 CRFs, 5 URFs) (Table 4). CRF02_AG was the most prevalent recombinant form accounting for 47.6% of all CRFs and 21.8% of all recombinants. It was common in Cameroon, Nigeria and Ghana. An intergroup recombinant, 02O was among the Cameroonian sequences.

All the countries in the Southern African region had no CRFs but a few URFs, while the countries in East Africa had more URFs than CRFs. In contrast, there were more CRFs than URFs in the countries in Central and West Africa (Table 4). Also, more recombinants were detected in our analysis compared to earlier analysis using partial sequences.

Table 5 shows regional differences in the distribution of some subtypes when partial and full-length sequences were used in the analysis. The detection rate of the subtypes A and G were less when full-length genome sequences are used in analysis compared with partial sequences, especially in regions where recombinants are common.

Table 1. Frequency of the HIV-1 groups and subtypes using full-length genome sequences. Figures from an earlier study (Hemelaar et. al are indicated).

HIV-1 variant	Number of full-length sequences	% of total HIV-1 subtypes	Hemelaar et al. (2006) (%)
Group M	903	96.7	
Subtype A	89	9.5	14%
Subtype B	3	0.3	N/S
Subtype C	446	47.8	56%
Subtype D	60	6.4	N/S
Subtype F	7	0.7	N/S
Subtype G	23	2.5	10%
Subtype H	1	0.1	N/S
Subtype J	1	0.1	N/S
Subtype K	2	0.2	N/S
Recombinants	271	29	16%
(CRF)	(124)	(13.3)	N/S
(URF)	(147)	(15.7)	N/S
Group N	10	1.1	
Group O	20	2.1	
Group P	1	0.1	
TOTAL	934	100	

N/S - Not stated.

DISCUSSION

Despite Africa bearing the brunt of HIV infection, there is limited information on the molecular epidemiology of HIV-1 due to the paucity and uneven availability of both partial and full-length genome sequences across the continent. Results from previous analyses (Hemelaar et al., 2006), had shown that the subtype A accounted for 21% of HIV-1 infections in West Africa, but in our analysis, it represented only 4.2% of sequences. Also, the subtype G previously observed to represent 35% of infections, accounted for only 16.9% of sequences. Though the subtype A was projected to represent 29% of infections in Nigeria, there were no available full-length or near full-length subtype A sequences from Nigeria. These analysed partial sequences in their survey, might really be part of CRF02_AG recombinants, but on their own appear as subtypes A or G. The earlier analysis had used sequences irrespective of the length, but weighted the distribution according to the number of HIV-infected people in each country. Disparities in subtype frequencies between our analysis and the Hemelaar study were also observed in some countries from the different sub-regions. This implies that the true HIV-1 subtype distribution might not have been captured using partial sequences.

Our analysis shows that recombinants (CRFs/URFs) constituted a substantial proportion of HIV-1 genotypes in sub-Saharan Africa. Our estimates of CRFs were higher than those obtained in a comprehensive audit of HIV distribution in 2004 (Hemelaar et al., 2006).

In that analysis, recombinants were projected to account for 42.6% of HIV-1 infections for West Africa, while we observed 71.8% for the same region. The differences in estimates might be due to the different approaches used, and also the timing because the Hemelaar study evaluated 2003/2004, while we analysed all available full-length sequences at the time of analysis. Our analysis sought to present the subtype distribution as based on full-length HIV-1 sequences available in the Los Alamos database. Whilst not a perfect approach, it presents the genetic diversity as determined by the unambiguity of full-length genome sequences.

There has been a consistent increase in the reporting of CRFs and URFs (Vidal et al., 2000; Nyombi et al., 2008). This is further buttressed by the fact that 21 new CRFs have been characterized between 2008 and 2013, implying that the already complex genetic diversity of HIV-1 is evolving further.

The increasing number of CRFs and their relative spread is also a reason for more full-length sequencing and analysis. This is important because the clinical implications of subtype variation with regards to recombinants are yet to be established. The spatial distribution of CRFs and URFs needs clarification as CRFs were common in Central and West Africa, while URFs were common in East Africa.

Our analysis further reveals the dearth of HIV-1 sequence information from sub-Saharan Africa, as there were only 934 full-length sequences from 27 countries having millions of people living with HIV. Hemelaar et al. (2011) in a later review had noted that the available

Table 2. Regional distribution of HIV-1 subtypes using full-length genome sequences.

Region	Country	*People living with HIV 2012	HIV* prevalence	A	B	C	D	F	G	H	J	K	CRF/URF	Group N	Group O	Group P	Total
Central Africa	Angola	250 000	2.3			1	1						1				3
	Cameroon	600 000	4.5	1			4	6	8		1	1	96	10	16	1	144
	CAR	N/A	N/A							1			3				4
	Chad	210 000	2.7				4										4
	DRC	480 000	1.1	4			4					1	18				27
	Gabon	41 000	4		1								2		2		5
East Africa	Djibouti	9 700	1.2			1											1
	Ethiopia	760 000	1.3			2											2
	Kenya	1 600 000	6.1	45		2	3		3				38				91
	Rwanda	240 000	2.9	8									2				10
	Somalia	29 000	0.5			1											1
	Tanzania	1 500 000	5.1	5		22	1						22				50
	Uganda	1 500 000	7.2	22			42						31				95
Southern Africa	Botswana	340 000	23			51							2				53
	Malawi	1 100 000	10.1			20											20
	South Africa	6 100 000	17.9	1	2	326	1						4				334
	Zambia	1 100 000	12.7			18							1				19
West Africa	Benin	72 000	1.1										1				1
	Gambia	14 000	1.3										3				3
	Ghana	240 000	1.4						2				18				20
	Guinea Bissau	41 000	3.9										2				2
	Ivory Coast	450 000	3.2										6				6
	Liberia	22 000	0.9										1				1
	Mali	100 000	0.9										1				1
	Niger	46 000	0.5										3				3
	Nigeria	3 400 000	3.1			1			10				10				21
	Senegal	43 000	0.5	3		1	1						6		2		13
Total		20 287 700		89	3	446	60	7	23	1	1	2	271	10	20	1	934

* UNAIDS Report on the Global AIDS epidemic 2013; DRC, Democratic Republic of Congo; CAR, Central African Republic; N/A Not available.

Table 3. Distribution of circulating and unique recombinant forms.

Region	Country	Circulating recombinant forms (CRFs)																					Unique Recombinant Forms (URFs)	
		1	2	6	9	10	11	13	16	18	21	22	25	26	27	30	32	36	37	45	49	Total	(Number of URFs indicated in brackets)	Total
Central Africa	Angola		1																			1		0
	Cameroon		36				7	6		1		10	3					2	2	1		68	0222 (6), 0213 (1), 0225 (1), 0102A (1), 0122F (1), 01ADF2 (1), 01A1G (1), 01F2 (1), 01F2G (1), 02A1U (1), 02B (1), 02F2G (1), 02G (1), 02O (1), A1F2 (1), A1G (1), A1H (1), AF2 (1), AF2G (1), AGU (1), AHJU (1), 22A1U (1), 22DU(1)	28
	CAR	3																				3		0
	DRC				1									4	2					2		9	26C (4), U (2), AGKU (1), A2G (1), 26CU (1)	9
	Gabon																			1		1	A1GHU (1)	1
East Africa	Kenya								1		3											4	AID (9), A1C (7), CD (4), A1CD (3), A1A2D (3), A2CD (2), A1CG (2), A1G (1), A1A2CD (1), AC (1), DG (1)	34
	Rwanda																					0	A1C (1), CD(1)	2
	Tanzania					3																3	A1C (12), A1D (4), CD (2), ACD (1)	19
	Uganda																					0	A1D (26), A1C (2), A1CD (1), A1G (1), CD (1)	31
Southern Africa	Botswana																					0	CD (1), A1GJ (1)	2
	South Africa																					0	A1CDGKU (1), A2C (1), AC (1), A1C (1)	4
	Zambia																					0	A2C (1)	1
West Africa	Benin																1					1		0
	Gambia																				3	3		0
	Ghana		9	1	1																	11	02D (2), 02A1 (2), 02A (1), 02G (1), 0206 (1)	7
	Guinea Bissau		2																			2		0
	Ivory Coast			1																		1	0209 (4), 02GK (1)	5
	Liberia		1																			1		0
	Mali			1																		1		0
	Niger															3						3		0
	Nigeria		8																			8	02A1 (1), A1G (1)	2
	Senegal		2	1	1																	4	02A1 (2)	2
Total		3	59	4	3	3	7	6	1	1	3	10	3	4	2	3	1	2	2	4	3	124		147

1, CRF01_AE; 2, CRF02_AG; 6, CRF06_cpx; 9, CRF09_cpx; 10, CRF10_CD; 11, CRF11_cpx; 13, CRF13_cpx; 16, CRF16_A2D; 18, CRF18_cpx; 21, CRF21_A2D; 22, CRF22_01A1; 25, CRF25_cpx; 26, CRF26_AU; 27, CRF27_cpx; 30, CRF30_0206;32, CRF32_06A1; 36, CRF36_cpx; 37, CRF37_cpx; 45, CRF45_cpx; 49, CRF49_cpx.

Table 4. Frequency of recombinant forms in some countries.

Country	% Recombinants	% CRFs	% URFs
Cameroon	66.7	47.2	19.4
DRC	66.7	33.3	33.3
Ghana	90	55	35
Kenya	41.8	4.4	37.4
Nigeria	47.6	38.1	9.5
Tanzania	44	6	38
Uganda	32.6	0	32.6

Table 5. Regional distribution of Recombinants. The proportion of regional sequences that are recombinants in this analysis and the Hemelaar et. Al. (2006) study are indicated.

Region	Country	CRFs	URFs	Total recombinants	Hemelaar et al. (2006) (%)
Central Africa	Angola	1	0	1	
	Cameroon	68	28	96	
	CAR	3	0	3	64.2%
	DRC	9	9	18	18.6
	Gabon	1	1	2	
	Regional Total	82	38	120	
East Africa	Kenya	4	34	38	
	Rwanda	0	2	2	
	Tanzania	3	19	22	37.2%
	Uganda	0	31	31	28.6
	Regional Total	7	86	93	
Southern Africa	Botswana	0	2	2	
	South Africa	0	4	4	1.2%
	Zambia	0	1	1	0.25
	Regional Total	0	7	7	
West Africa	Benin	1	0	1	
	Gambia	3	0	3	
	Ghana	11	7	18	
	Guinea Bissau	2	0	2	
	Ivory Coast	1	5	6	
	Liberia	1	0	1	71.8%
	Mali	1	0	1	42.6
	Niger	3	0	3	
	Nigeria	8	2	10	
	Senegal	4	2	6	
	Regional Total	35	16	51	
Total		124	147	271	

sequences were not representative of the HIV-1 distribution in the countries of origin, and that some countries harbouring large numbers of infected individuals with high subtype diversity had a small amount of HIV data. This is particularly true of a country like Nigeria that has a relatively high HIV-1 burden, but has only 21 available full-length HIV-1 sequences.

In a review of selected studies that have documented HIV subtype diversity in East, West and Southern Africa, only two studies were observed to have used near full-

length and full-length genome sequences (Lihana et al., 2012). There is thus a need for more full-length sequencing. Cost and lack of requisite equipment and manpower are probably responsible for the gross under-representation of sub-Saharan HIV-1 full-length sequences. Worrisome is the fact that of over 2148 high-throughput sequencing machines in the world, there are only 17 in Africa (3 in Kenya, 14 in South Africa) (http://omicsmaps.com/). Due to cost or other limitations, most studies in Africa are limited to sequencing of partial HIV genomes, and even these studies are identifying recombination within partial gene sequences (Kiwelu et al., 2013), so the full extent of genetic variability and recombination could be obscured unless the full-length genome is sequenced and analysed. Studies using partial sequences, have shown an increase in the detection of URFs and drug resistant viruses in sub-Saharan Africa (Ragupathy et al., 2011; Jacobs et al., 2014). Analysis of full-length sequences could possibly lead to greater identification of these recombinant forms.

Analysis of full-length sequences could also help in the accurate identification of low frequency viral variants (Henn et al., 2012) and the use of multiple genes rather than single gene to identify HIV-1 subtypes can reduce the chances of false identification (Neogi et al., 2012). The fact that the full-length subtypes E and I isolates were never found, and have now been re-designated as circulating recombinant forms, CRF01_AE and CRF04_cpx respectively (Carr et al., 1996; Gao et al., 1998; Paraskevis et al., 2001), and the suggestion that the subtype G, was actually a recombinant, whose parental subtype included the CRF02_AG (Abecasis et al., 2007), justify the calls for subtype classification to be based only on analysis of full-length or near full-length genomes.

It follows that analysis of partial HIV-1 sequences could be misinterpreted and may not reveal the true picture of HIV-1 biology and pathogenesis. Therefore, there is the need to know the current incidence/distribution of HIV-1 and also the need to expand the subtype database as these may impact on diagnosis, therapy and vaccine design. Full-length sequences are probably the most accurate representation of HIV genetic diversity.

Conclusion

This analysis brings to light the need for more sequences of full-length genomes from the sub-Saharan Africa region. This is a herculean task because even partial sequences are difficult to come by in most countries in the region. It will require an understanding of the importance of sequencing, commitment from govern-ments within the region and continuous hard work from scientists to achieve this objective. The periodic monitoring of HIV variants could help determine the extent of virus evolution.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

Abebe A, Demissie D, Goudsmit J, Brouwer M, Kuiken CL, Pollakis G Schuitemaker H, Fontanet AL, Rinke de Wit TF (1999). HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS. 13:1305-1311.

Abecasis AB, Lemey P, Vidal N, de Oliveira T, Peeters M, Camacho R et al (2007). Recombination Confounds the Early Evolutionary History of Human Immunodeficiency Virus Type 1: Subtype G Is a Circulating Recombinant Form. J. Virol. 81:8543-8551.

Alaeus A, Lidman K, Bjorkman A, Giesecke J, Albert J (1999). Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. AIDS 13:901-907.

Baeten JM, Chohan B, Lavreys L, Chohan V, McClelland RS, Certain L et al (2007). HIV-1 Subtype D Infection Is Associated with Faster Disease Progression than Subtype A in Spite of Similar Plasma HIV-1 Loads. J. Infect. Dis. 195:1177-1180.

Bjorndal A, Deng HK, Jansson M, Fiore JR, Colognesi C, Karlsson A et al (1997). Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. J. Virol. 71:7478-7487.

Blackard JT, Renjifo B, Fawzi W, Hertzmark E, Msamanga G, Mwakagile D, Hunter D, Spiegelman D, Sharghi N, Kagoma C, Essex M (2001). HIV-1 LTR Subtype and Perinatal Transmission. Virology 287:261-265.

Boyer JC, Bebenek K, Kunkel TA (1992). Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. Proc. Natl. Acad. Sci. 89:6919-6923.

Carr JK, Salminen MO, Koch C, Gotte D, Artenstein AW, Hegerich PA et al (1996). Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. J. Virol. 70:5935-5943.

Cleghorn FR, Jack N, Carr JK, Edwards J, Mahabir B, Sill A et al (2000). A distinctive clade B HIV type 1 is heterosexually transmitted in Trinidad and Tobago. Proc. Natl. Acad. Sci. 97:10532-10537.

Connell BJ, Michler K, Capovilla A, Venter, WDF, Stevens WS, Papathanasopoulos MA et al (2008). Emergence of X4 usage among HIV-1 subtype C: evidence for an evolving epidemic in South Africa. AIDS 22:896-899

Domingo E, Menendez-Arias L, Holland JJ (1997). RNA virus fitness. Rev.Med. Virol. 7:87-96.

Easterbrook P, Smith M, Mullen J, O'Shea S, Chrystie I, de Ruiter A et al (2010). Impact of HIV-1 viral subtype on disease progression and response to antiretroviral therapy. J. Intl. AIDS Soc. 13:4.

Esbjornsson J, Mansson F, Martinez-Arias W, Vincic E, Biague A, da Silva Z et al (2010). Frequent CXCR4 tropism of HIV-1 subtype A and CRF02_AG during late-stage disease - indication of an evolving epidemic in West Africa. Retrovirology 7:23.

Gao F, Robertson DL, Carruthers CD, Li YY, Bailes E, Kostrikis LG et al (1998). An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). J. Virol. 72:10234-10241.

Gao F, Vidal N, Li YY, Trask SA, Chen YL, Kostrikis LG et al. (2001). Evidence of two distinct subsubtypes within the HIV-1 subtype A radiation. AIDS Res. Hum. Retroviruses 17:675-688.

Hemelaar J, Gouws E, Ghys PD, Osmanov S (2006). Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 20: W13-W23.

Hemelaar J, Gouws E, Ghys PD, Osmanov S, WHO-UNAIDS Network for HIV Isolation and Characterisation (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS 25:679-689.

Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR (2012). Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. PLoS Pathog 8:e1002529.

Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature. 373:123.

Hu W, Temin HM (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: Pseudodiploidy and high rate of genetic recombination. Proc. Natl. Acad. Sci. 87:1556-1560.

Jacobs GB, Wilkinson E, Isaacs S, Spies G, de Oliveira T, Seedat S et al (2014). HIV-1 Subtypes B and C Unique Recombinant Forms (URFs) and Transmitted Drug Resistance Identified in the Western Cape Province, South Africa. PLoS ONE 9:e90845.

Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP (2000). High rate of recombination throughout the human immunodeficiency virus type 1 genome. J. Virol. 74:1234-1240.

Johnston ER, Zijenah LS, Mutetwa S, Kantor R, Kittinunvorakoon C, Katzenstein DA (2003). High frequency of syncytium-inducing and CXCR4-tropic viruses among human immunodeficiency virus type 1 subtype C-infected patients receiving antiretroviral Treatment. J. Virol. 77:7682-7688.

Kaleebu P, Nankya IL, Yirrell DL, Shafer LA, Kyosiimire-Lugemwa J, Lule DB et al (2007). Relation Between Chemokine Receptor Use, Disease Stage, and HIV-1 Subtypes A and D: Results From a Rural Ugandan Cohort. JAIDS 45:28-33

Kaleebu P, Ross A, Morgan D, Yirrell D, Oram J, Rutebemberwa A et al (2001). Relationship between HIV-1 Env subtypes A and D and disease progression in a rural Ugandan cohort. AIDS. 15:293-299.

Kanki PJ, Hamel DJ, Sankale JL, Hsieh CC, Thior I, Barin F et al (1999). Human immunodeficiency virus type 1 subtypes differ in disease progression. J. Infect. Dis. 179:68-73.

Kiwelu IE, Novitsky V, Margolin L, Baca J, Manongi R, Sam N et al (2013). Frequent Intra-Subtype Recombination among HIV-1 Circulating in Tanzania. PLoS ONE 8:e71131.

Lihana RW, Ssemwanga D, Abimiku A, Ndembi N (2012). Update on HIV diversity in Africa: A decade in review. AIDS Rev. 14:83-100.

Meloni ST, Kim B, Sankale J-L, Hamel DJ, Tovanabutra S, Mboup S et al (2004). Distinct Human Immunodeficiency Virus Type 1 Subtype A Virus Circulating in West Africa: Sub-Subtype A3. J. Virol. 78: 12438-12445.

Meyerhans A, Cheynier R, Albert J, Seth M, Kwok S, Sninsky J et al (1989). Temporal fluctuations in HIV Quasispecies *in vivo* are not reflected by sequential HIV isolations. Cell 58:901-910.

Neogi U, Gupta S, Shet A, De Costa A, Laishram RL, Wanchu A et al (2012). Emergence of unique recombinant forms (URFs) in Indian HIV-1 epidemic: data from nationwide clinical cohort between 2007 and 2011. Retrovirology. 9(Suppl 2):151.

Nyombi BM, Kristiansen KI, Bjune G, Müller F, Holm-Hansen C (2008). Diversity of Human immunodeficiency virus type 1 subtypes in Kagera and Kilimanjaro regions, Tanzania. AIDS Res. Hum. Retroviruses. 24:761-769.

Paraskevis D, Magiorkinis M, Vandamme AM, Kostrikis LG, Hatzakis A (2001). Re-analysis of human immunodeficiency virus type 1 isolates from Cyprus and Greece, initially designated 'subtype I', reveals a unique complex A/G/H/K/? mosaic pattern. J. Gen. Virol. 82:575-580.

Peeters M, Vincent R, Perret J-L, Lasky M, Patrel D, Liegeois F et al (1999). Evidence for differences in MT2 cell tropism according to genetic subtypes of HIV-1: Syncytium-inducing variants seem rare among subtype C HIV-1 viruses. J. Acquir. Immune. Defic. Syndr. Hum. Retrovirol. 20:115-121.

Plantier J, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemee V et al (2009). A new human immunodeficiency virus derived from gorillas. Nat. Med. 15:871.

Ragupathy V, Zhao J, Wood O, Tang S, Lee S, Nyambi P et al (2011). Identification of new, emerging HIV-1 unique recombinant forms and drug resistant viruses circulating in Cameroon. Virol. J. 8(185).

Renjifo B, Chung M, Gilbert P, Mwakagile D, Msamanga G, Fawzi W, Essex M (2003). In-utero transmission of quasispecies among human immunodeficiency virus type I genotypes. Virology 307:278-282.

Renjifo B, Fawzi W, Mwakagile D, Hunter D, Msamanga G, Spiegelman D et al (2001). Differences in perinatal transmission among human immunodeficiency virus type 1 genotypes. J. Hum. Virol. 4:16-25.

Roberts JD, Bebenek K, Kunkel TA (1988). The accuracy of reverse transcriptase from HIV-1. Science 242:1171-1173.

Sampathkumar R, Shadabi E, Luo M (2012). Interplay between HIV-1 and Host Genetic Variation: A Snapshot into Its Impact on AIDS and Therapy Response. Advances Virol. 2012:508967(doi:10.1155/2012/508967).

Triques K, Bourgeois A, Saragosti S, Vidal N, Mpoudi-Ngole E, Nzilambi N et al (1999). High diversity of HIV-1 subtype F strains in Central Africa. Virology 259:99-109.

van Harmelen J, Williamson C, Kim B, Morris L, Carr J, Karim SSA, McCutchan F (2001). Characterization of full-length HIV type 1 subtype C sequences from South Africa. AIDS Res. Hum. Retroviruses 17:1527-1531.

van Harmelen J, Wood R, Lambrick M, Rybicki EP, Williamson AL, Williamson C (1997). An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. AIDS 11:81-87.

Vasan A, Renjifo B, Hertzmark E, Chaplin B, Msamanga G, Essex M et al (2006). Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. Clin. Infect. Dis. 42:843-852.

Vidal N, Bazepeo SE, Mulanga C, Delaporte E, Peeters M (2009). Genetic Characterization of Eight Full-Length HIV Type 1 Genomes from the Democratic Republic of Congo (DRC) Reveal a New Subsubtype, A5, in the A Radiation That Predominates in the Recombinant Structure of CRF26_A5U. AIDS Res. Human Retroviruses 25:823-832.

Vidal N, Mulanga C, Bazepeo SE, Lepira F, Delaporte E, Peeters M et al (2006). Identification and molecular characterization of subsubtype A4 in Central Africa. AIDS Res. Human Retroviruses 22:182-187.

Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W et al (2000). Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. J. Virol. 74:10498- 10507.

Yebra G, de Mulder M, Martín L, Rodríguez C, Labarga P, Viciana I et al (2012). Most HIV Type 1 Non-B Infections in the Spanish Cohort of Antiretroviral Treatment-Naïve HIV-Infected Patients (CoRIS) Are Due to Recombinant Viruses. J.Clin. Microbiol. 50:407-413.

Zhuang JL, Jetzt AE, Sun GL, Yu H, Klarmann G, Ron Y et al (2002). Human immunodeficiency virus type 1 recombination: Rate, fidelity, and putative hot spots. J. Virol. 76:11273-11282.

Full Length Research Paper

Chemical composition and larvicidal activity of *Zanthoxylum gilletii* essential oil against *Anopheles gambiae*

Ombito O. Japheth[1]*, Matasyoh C. Josphat[1] and Vulule M. John[2]

[1]Department of Chemistry, Egerton University, P. O. Box 536, Egerton - 20115 Kenya.
[2]Department of Entomology, KEMRI, P. O. Box 1578, Kisumu - 40100 Kenya.

Malaria is a serious health problem in many African countries. The *Anopheles gambiae* mosquito which is the major vector for this disease has developed resistance against synthetic pyrethroids which are the main stay of insecticide treated bed nets. The development of insecticide resistance and side effects associated with synthetic pesticides has triggered intense research efforts towards natural products (for vector control) such as essential oils because of their efficacy and safety. In this study, larvicidal potential of essential oil from *Zanthoxylum gilletii* was evaluated against malaria vector mosquito, *A. gambiae*. The essential oil was extracted by hydro-distillation, and its chemical compositions determined by gas chromatography mass spectrometry. The oil was dominated by sesquiterpenes and monoterpenes which accounted for 38.30 and 34.00%, respectively. The oil showed good activity against *A. gambiae* and recorded LC_{50} and LC_{90} values of 57.73 and 140.24 $\times 10^{-3}$ mg/ml, respectively. The results obtained show that the essential oil isolated from *Z. gilletii* is a promising mosquito larvicide.

Key words: Malaria, *Zanthoxylum gilletii*, essential oil, *Anopheles gambiae*, larvicidal activity

INTRODUCTION

Mosquitoes are known vectors of various diseases which are life threatening. *Anopheles gambiae* mosquitoes are known to transmit malaria (Cheng et al., 2003; Das and Ansari, 2003; Magalhaes et al., 2010). According to the latest WHO estimates, there were approximately 219 million cases of malaria globally in 2010 and 660,000 fatal cases: approximately 90% of these fatal cases occurred in Africa (WHO, 2012).

Currently there is no effective available vaccine for malaria (Matasyoh et al., 2008). Among the efforts that have been made in recent decades in seeking to reduce

mosquito bites and transmission of malaria include the use of insecticide treated nets (ITNs) and larviciding. The insecticide treated nets rely solely on pyrethroids to enhance their protective utility (Chavasse et al., 1999). Larviciding has the greatest control impact on mosquito populations because the larvae are concentrated, immobile and accessible, and it employs the use of synthetic insecticides (Tiwary et al., 2007). However, the overreliance on these synthetic chemicals for mosquito control has resulted in the development of insecticide resistance over time (Hemingway and Ranson, 2000). The

*Corresponding author. E-mail: jeffombito@gmail.com. Tel: +254 729 020575.

spread in resistance to majority of present synthetic insecticides by *A. gambiae* and the environmental pollution coupled with the safely risks for both human and domestic animals posed by these insecticides has highlighted the need for novel strategies for control of *A. gambiae* (Cheng et al., 2009b). Hence, there is a renewed interest in the exploration and use of plant products with insecticidal properties for mosquito control.

Aromatic plants and their essential oils are very important sources of many compounds that are used for different applications (Abduelrahman et al., 2009). Essential oils are simply volatile fractions obtained by either steam or water distillation of medicinal and aromatic plants (Rabha et al., 2012). Essential oils have received considerable renewed attention as potent bioactive compounds against various species of mosquitoes. They are potentially suitable for application in larval control management because they constitute a rich source of bioactive compounds that are effective and naturally biodegradable into non-toxic products (Lucia et al., 2007; Cheng et al., 2008; Cheng et al., 2009a).

Zanthoxylum gilletii is a tropical rainforest species, distributed between altitudes ranging from 900 to 2400 m. It is a valued forest tree that grows naturally but planted in Western Kenya for timber and medicinal properties (Kokwaro et al., 1976). The Luhya community; that is, a major habitat of this region uses the bark of *Z. gilletii* in traditional anti-malaria preparations (Nyunja et al., 2009).

The present study attempted to investigate the larvicidal efficacy of essential oils derived from *Z. gilletii* leaves against the medically important malaria vector mosquito, *A. gambiae* with the purpose of identifying effective indigenous bio-products to control the vector of mosquito-borne diseases, particularly in cases where the vector's susceptibility to conventional synthetics is decreasing.

MATERIALS AND METHODS

Sample collection

The leaves of *Z. gilletii* were collected from Kakamega forest, a tropical rain forest in Kenya which stretches from 0° 10' to 0° 21' N and longitude 34° 44' to 34° 58' E and an altitude of 1524 m above the sea level. The leaves were identified with the help of a taxonomist. Voucher specimens were deposited at the Department of Biological Sciences, Egerton University, Kenya.

Extraction of essential oil

Fresh leaves of *Z. gilletii* were cut into pieces less than 2 × 2 cm within 12 h after collection and 1000 g hydro-distilled in a modified type-Clevenger apparatus for 4 h. The essential oil obtained was dried over anhydrous Na_2SO_4 and kept in glass vial under refrigeration at 4°C.

Essential oil analysis

Samples of essential oils were diluted in methyl-t-butyl ether (MTBE) (1:100) and analyzed on an Agilent GC-MSD apparatus equipped with an Rtx-5SIL MS ('Restek') (30 m × 0.25 mm, 0.25 µm

film thickness) fused-silica capillary column. Helium (at 0.8 ml/min) was used as a carrier gas. Samples were injected in the split mode at a ratio of 1:10 to 1:100. The injector was kept at 250°C and the transfer line at 280°C. The column was maintained at 50°C for 2 min and then programmed to 260°C at 5°C/min and held for 10 min at 260°C. The MS was operated in the electron impact ionization (EI) mode at 70 eV, in m/z ranging from 42 to 350. The identification of the compounds was performed by comparing their retention indices and mass spectra with those found in literature (Adams, 2007) and supplemented by Wiley 7N.l, HPCH 1607.L and FLAVORS.L GC-MS libraries. The relative proportions of the essential oil constituents are expressed as percentages obtained by peak area normalization, all relative response factors being taken as one.

Mosquito larvicidal bioassays

The larval toxicity tests were carried out following the standard World Health Organization larval bioassay method (WHO, 2005), with slight modifications. Since oil does not dissolve in water, it was first solubilized in dimethyl-sulphoxide (DMSO, analytical reagent, Lobarchemi) and diluted with spring river water to make a stock solution of 1000 mg/L. Serial dilutions of the stock solution were done at different concentrations which included 500, 250, 200, 150, 125, 100, 62.5, 55, 45, 40, 31.25, 15.6 and 7.8×10^{-3} mg/ml. The concentration of DMSO was kept below 1% since at this level it does not affect larval mortality. The bioassays were conducted at the Kenya Medical Research Institute (KEMRI), Centre for Global Health Research (CGHR), Kisumu, Kenya, where the insects were reared in plastic and enamel trays in spring river water. They were maintained and all experiments were carried out at 26 ± 3°C and the humidity ranged between 70 and 75%. The bioassays were performed with third instar larvae of *A. gambiae* and carried out in triplicate using 20 larvae for each replicate assay. The replicates were run simultaneously yielding a final total of 60 larvae for each concentration. The larvae were collected by direct pipetting from the enamel trays and transferred to 25 ml disposable plastic cups containing 10 ml of test solution and fed on tetramin fish feed during all testing. Mortality and survival was established after 24 h of exposure. Larvae were considered dead if they were unrousable within a period of time, even when gently prodded with a micro-pipette. The dead larvae in the three replicates were combined and expressed as the percentage mortality for each concentration. The negative control was 1% DMSO in spring river water while the positive control was the pyrethrum based larvicide, pylarvex.

Statistical analysis

The average larval mortality data were subjected to probit analysis for calculating LC_{50} and LC_{90} at 95% fiducial limits of upper confidence limit and lower confidence limit (Finney, 1971) using IBM SPSS software version 15.

RESULTS AND DISCUSSION

The essential oil of *Z. gilletii* was dominated by monoterpenes and sesquiterpenes which accounted for 34.00 and 38.30%, respectively. The major monoterpene components included γ-terpinene (10.62%), β-myrcene (5.16%), sabinene (4.89%), β-ocimene (3.12%) and camphene (2.56%). The main sesquiterpene components were *trans*-caryophyllene (9.82%), caryophyllene oxide (4.4%), α-cadinol (2.71%), 1, 1, 4, 8-tetramethyl-4, 7, 10-cycloundecatriene (2.62%), δ-cadinene (2.52%) and τ-cadinol (2.29%) (Table 1).

The essential oil of *Z. gilletii* was active against third

Table 1. Identified compounds of the essential oil from *Zanthoxylum gilletii.*

Compound name	Retention time (min)	Concentration (%)
Monoterpenes		
γ-terpinene	6.93	10.62
β- myrcene	8.54	5.16
Sabinene	7.96	4.89
β-Ocimene	10.19	3.12
Camphene	7.25	2.56
Alloocimene	12.37	1.35
Bornyl acetate	16.77	1.16
3,7-dimethyl-1,6-octadien-3-ol	11.53	0.84
n- Decanal	14.55	0.58
Terpine-4-ol	13.73	0.55
Cis-epoxyocimene	12.73	0.44
2-methyl-2-phenylpropanal	15.53	0.37
Tricyclene	8.81	0.32
(2-methylpropyl)-benzene	16.01	0.16
trans-Sabinene hydrate	10.61	0.13
trans-(+)-carveol	14.95	0.12
Total		34.00
Sesquiterpenes		
trans-Caryophyllene	20.46	9.82
Caryophyllene oxide	24.39	4.40
α-Cadinol	25.99	2.71
1,1,4,8-tetramethyl-4,7,10-cycloundecatriene	21.26	2.62
δ-Cadinene	22.84	2.52
т -Cadinol	25.67	2.29
β-Cubebene	19.56	1.70
2-isopropyl-5-methyl-9-methylene-bicyclo [4.4.0] dec-1-ene	22.26	1.43
β-Selinene	21.99	1.35
Germacrene D	20.99	0.87
1a,2,3,4,4a,5,6,7b-octahydro-1,1,4,7-tetramethyl-1H-cycloprop (e) azulene	23.21	0.84
2,6,6-trimethyl-5-(3-methyl-2-butyl)-1-cyclohexene-1-methanol	23.87	0.77
α-Copaene	19.15	0.58
3-thujopsanone	30.22	0.39
Juniper camphor	26.86	0.31
β-oplopenone	29.05	0.30
Cis-nerolidol	33.43	0.29
α-Cubebene	18.46	0.23
Total		38.30
Diterpenes		
Phytol	35.02	1.51
5-(decahydro-5,5,8a-trimethyl-2-methylene-1-naphthalenyl)-3-methyl-2-pentenoic acid	33.96	0.28
2,6,10-trimethyl-13-(1-methylethenyl)-2,5,9-cyclotetradectrien-1-ol	38.53	0.04
n-eicosane	41.32	0.04
Total		8.50

Table 1. Contd.

Others		
2-Undecanone	17.44	3.64
Cryptone	14.06	1.00
4-ethenyl-cyclohexenemethanol	23.55	0.98
Octahydro-4a-methyl-7-(1-methylethyl)-2(1H)-naphthalenone	30.81	0.84
8-dodecenol	27.13	0.30
Palmitic acid	32.24	0.24
Nonadecane	44.30	0.24
Nonacosane	47.95	0.16
(1S,2S,5R)-(+)-4-isopropyl-7-methyl-1-oxaspiro(2,5) octane	50.25	0.03
Total		19.10
Total percentages		99.90

Table 2. Larvicidal activity of essential oil against third instar larvae of *An. gambiae* after 24 h of exposure.

Concentration ($\times 10^{-3}$ mg/mL)	% Mortality ± SD	LC_{50} ($\times 10^{-3}$ mg/ml)	LC_{90} ($\times 10^{-3}$ mg/ml)
7.80	0.00 ± 0.00		
15.60	1.67 ± 2.89		
31.25	3.33 ± 2.89		
40.00	13.33 ± 5.77		
45.00	33.33 ± 5.77		
55.00	46.67 ± 2.89		
62.50	68.33 ± 5.77		
100.00	76.67 ± 2.89	57.73 (45.40-73.05)	140.24 (105.73-217.05)
125.00	80.00 ± 10.00		
150.00	85.00 ± 8.66		
200.00	88.33 ± 2.89		
250.00	96.67 ± 5.77		
500.00	100.00 ± 0.00		
1000.00	100.00 ± 0.00		
Pylarvex (0.1 mg/mL)[X]	100.00 ± 0.00		
Spring water + DMSO[Y]	0.00 ± 0.00		

[X]Positive control, [Y]Negative control.

instar larvae of *A. gambiae* with LC_{50} and LC_{90} values of 57.73 and 140.24 $\times 10^{-3}$ mg/ml respectively (Table 2). The negative control showed no activity against third instar larvae of *A. gambiae*. At a concentration of 100×10^{-3} mg/ml, the positive control pyrethrin EC 0.5% w/v (Pylarvex™) recorded 100 % larval mortality (Table 2).

Compared to essential oils from other plants, the essential oil of *Z. gilletii* exhibits significant activity against malaria vector *A. gambiae*. Reports from earlier studies indicate that larvicidal activity of essential oil from *F. angolensis* (Rutaceae) against third instar larvae of *A. gambiae* exhibited LC_{50} and LC_{90} values of 83.7 and 324.0 mg/L, respectively (Mudalungu et al., 2013). Kweka et al. (2012) reported larvicidal activity of *P. amboinicus*

essential oil against late third instar larvae of *A. gambiae* and observed LC_{50} and LC_{90} values of 67.53 and 107.60 ppm, respectively. The essential oil of *C. citratus* was observed to have an LC_{50} of 69 ppm against *Aedes aegypti* larvae (Cavalcanti et al., 2004). The same oil was reported to possess larvicidal activity against *Culex quinquefasciatus* larvae with LC_{50} of 165.7 ppm (Pushpanathan et al., 2008) (Table 3).

γ-Terpinene, the major component of *Z. gilletii* essential oil has been found to possess larvicidal activity against three mosquito species (Cheng et al., 2009c; Zhu and Tian., 2011). This compound had LC_{50} values of 26.8 µg/ml, 22.8 µg/ml and 29.21 mg/L against *A. aegypti*, *Aedes albopictus* and *A. anthropophagus* larvae,

Table 3. Comparison of relative toxicity of essential oil from *Zanthoxylum gilletii* with three previously isolated essential oils tested against *A. gambiae*, *A. aegypti* and *C. Quinquefasciatus*.

Essential oil	Mosquito species	LC_{50} (mg/ml)	LC_{90} (mg/ml)
Zanthoxylum gilletii	*Anopheles gambiae*	57.73	140.24
Plectranthus amboinicus	*Anopheles gambiae*	67.53	107.60
Cymbopogon citrates	*Aedes aegypti*	69.00	
	Culex quinquefasciatus	165.70	
Fagaropsis angolensis	*Anopheles gambiae*	83.70	324.00

respectively. It recorded an LC_{90} value of 63.1mg/L against *A. anthropophagus*. In a previous study, the same compound was isolated from the oil fractions of *Cymbopogon nardus* and had an excellent effect against third instar larvae of *C. quinquefasciatus* in 24 h, with LC_{50} value of 0.8 mg/L (Ranaweera and Dayananda, 1996).

Sabinene which was also in appreciable amount in the oil has been reported to exhibit larvicidal activity against third instar larvae of *C. quinquefasciatus*, *A. aegypti* and *Anopheles stephensi* (Govindarajan, 2010). The LC_{50} values recorded against the three mosquito species were 25.01, 21.20 and 19.67 ppm, respectively. The compound also recorded LC_{90} values of 45.15 ppm against *C. quinquefasciatus*, 39.22 ppm against *A. aegypti* and 36.45 ppm against *A. stephensi*. Sabinene was also isolated from the essential oil of *Clausena dentata* and found to be active against *Spodoptera litura* with LC_{50} and LC_{90} values of 21.42 and 40.39 ppm respectively (Krishnappa et al., 2010). Another monoterpene also present in appreciable amount in the oil was β-myrcene. Previous studies have documented the activity of β-Myrcene against the larvae of *A. aegypti* and *A. albopictus* (Cheng et al., 2009c). The LC_{50} values recorded were 27.9 and 23.5 µg/ml respectively.

Trans-Caryophyllene, which occurs in appreciable amounts in this oil, is also reported to show activity against *A. aegypti* larvae with LC_{50} of 104 ppm (Morais et al., 2006). Its oxygenated form caryophyllene oxide is known to exhibit larvicidal activity against the fourth instar larvae of *A. anthropophagus* (Zhu and Tian, 2013) with LC_{50} and LC_{90} values of 49.46 and 115.38 mg/L, respectively. α-Cadinol although in small quantity in *Z. gilletii* oil, is known to possess larvicidal activity against *A. aegypti* with LC_{50} value of 76.1 ppm (Chun et al., 2008). Germacrene D is known to be effective against larvae of *A. aegypti* and *A. stephensi* (Kiran et al., 2006). This sesquiterpene hydrocarbon, isolated from the essential oils of *Chloroxylon swietenia*, had LC_{50} values of 63.6 and 59.5 µg/ml against *A. aegypti* and *A. stephensi* respectively. The LC_{90} values recorded for this compound were 100.7 µg/ml against *A. aegypti* and 96.4 µg/ml against *A. stephensi*.

The high larvicidal activity of *Z. gilletii* can therefore be attributed to the presence of γ-terpinene, β-myrcene, sabinene, *trans*-caryophyllene, caryophyllene oxide, α-

cadinol and germacrene D which have been documented to possess larvicidal activity against different species of mosquito.

Conclusion

Plants are rich source of bioactive organic chemicals and offer an advantage over synthetic pesticides as they are less toxic, less prone to development of resistance, and easily biodegradable. The findings of this study show that the essential oil isolated from *Z. gilletii* holds great promise as potential mosquito larvicides. Furthermore these outcomes could be useful in the search for newer more selective, biodegradable and natural larvicidal compounds. These findings also offer an opportunity for developing alternatives to inorganic insecticides.

REFERENCES

Abduelrahman A, Elhussein S, Osman N, Nour A (2009). Morphological Variability and Chemical Composition of Essential Oils from nineteen Accession of Basil (*Ocimum basilicum* L) Growing in Sudan. Int. J. Chem. Techn. 1(1): 1-10.

Adams RP (2007). Identification of essential oil components by Gas Chromatography/ Mass Spectrometry 4th ed. Allured publishing corporation: Carol Stream. Illinois. USA.

Cavalcanti ESB, de Morais SM, Lima MAA, Santana EWP (2004). Larvicidal Activity of Essential Oils from Brazilian Plants against *Aedes aegypti* L. Mem Inst Oswaldo Cruz, Rio de Janeiro. 99(5):541-544.

Chavasse DC, Shier RP, Murphy OA, Huttly SRA, Cousens SN, Akhtar T (1999). Impact of fly control on childhood diarrhoea in Pakistan. Lancet. 353: 22-25.

Cheng SS, Chang HT, Chang ST,Tsai KH, Chen WJ (2003). Biactivity of selected plant essential oils against the yellow fever mosquito *Aedes aegypti* larvae. Bioresour. Technol. 89: 99-102.

Cheng SS, Huang CG, Chen WJ, Kuo YH, Chang ST (2008). Larvicidal activity of tectoquinone isolation from red heartwood-type *Cryptomeria japonica* against two mosquito species. Bioresour. Technol. 99: 3617-3622.

Cheng SS, Chua MT, Chang EH, Huang CG, Chen WJ, Chang ST (2009a). Variation in insecticidal activity and chemical composition from leaf essential oils of *Cryptoemeria japonica* at different ages. Bioresour. Technol. 100:465-470.

Cheng SS, Huang CG, Cheng YJ, Yu JJ, Chen WJ, Chang ST (2009b). Chemical Compositon and larvicidal activities of leaf essential from eucalyptus species. Bioresour. Technol. 100:452-456.

Cheng, SS, Chang, HT, Lin, CY, Chen, PS, Huang, CG, Cheng WJ, Chang ST (2009c). Insecticidal activities of leaf and twig essential oils from *Clausena excavata* against *Aedes aegypti* and *Aedes albopictus* larvae. Pest Manage. Sci. 65:339-343.

Chun YL, Guan YK, Chin GH, Shang TC, Wei JC, Hui TC (2008). Mosquito larvicidal activities of leaf essential oils from common *Michelia* species in Taiwan. Q. J. C. For. 41(4):559-567.

Das MK, Ansari MA (2003). Evalution of repellent action of *Cymbopopogan martinii martinii* stapf var sofia oil against *Anopheles sandiacus* in tribal villages of Car Nicobar Island Andamana and Nicobar Island India. J. Vect. Borne. Di. 40: 101-104.

Finney DJ (1971). Probit Analysis. 3rd ed. London: Cambridge University Press. pp. 38-39.

Govindarajan M (2010). Chemical composition and larvicidal activity of leaf essential oil from *Clausena anisata* (Willd.) Hook. f. ex Benth (Rutaceae) against three mosquito species. Asian Pac. J. Trop. Med. pp. 874-877.

Hemingway J, Ranson H (2000). Insecticide Resistance in Insect Vectors of Human Diseases. Annu. Rev. Entomol. 45: 371-391.

Kiran SR, Bhavani K, Devi SP, Rao BRR, Reddy KJ (2006). Composition and larvicidal activity of *Chloroxylon swietenia* DC against *Aedes aegypti* and *Anopheles stephensi*. Bioresour. Technol. 97:2481-2484.

Kokwaro JO (1976). Medicinal plants of East Africa. East African Literature Bureau.

Krishnappa K, Elumai K, Govindarajan AAM, Mathivanan T (2010). Chemical composition and larvicidal and ovicidal activity of essential oil from *Clausena dentata* against Armyworm, *Spodoptera litura* (FAB) (Leidoptera: Noctuidae). IJRSR. 8:188-203.

Kweka JE, Senthikumar A, Venkatesalu V (2012). Toxicity of essential oil from Indian borage on the larvae of African malaria vector mosquito, *Anopheles gambiae*. Parasites Vectors. 5:277.

Lucia A, Audino GA, Seccacini E, Licastro S, Zerba E, Masuh H (2007). Larvicidal effect of *Eucalyptus grandis* essential oil and terpentine and their major components against Aedes aegypti larvae. J. Am. Mosq. Contr. Assoc. 3:299-303.

Magalhaes L, Lima M, Marques M, Facanali R, Pinto A, Tadei W (2010). Chemical composition and larvicidal activity against *Aedes aegypti* larvae of essential oils from four *Guarea* species. Molecule 15:5734-5741.

Matasyoh JC, Wathuta EM, Kariuki ST, Chepkorir R, Kavulani J (2008). *Aloe* plant extracts as alternative larvicides for mosquito control. Afr. J. Biotechnol. 7:912-915.

Morais SM, Cavalcanti ESB, Bertini LM, Oliveria CLL, Rodrigues JRB, Cardoso JHL (2006). Larvicidal activityof essential oils from Brazilian Croton species against *Aedes aegypti* L. J. Am. Mosq. Contr. Assoc. 22:161-4.

Mudalungu CM, Matasyoh JC, Vulule JM, Chepkorir R (2013). Larvicidal compounds from *Fagaropsis angolensis* leaves against malaria vector (*Anopheles gambiae*). Int. J. Malaria. Res. Rev. 1(1):1-7.

Nyunja ARC, Onyango JC, Erwin B (2009). The Kakamega Forest Medicinal Plant Resources and their Utilization by the Adjacent Luhya Community. Int. J. Trop. Med. 3:82-90.

Pushpanathan T, Jebanesan A, Govindarajan M (2008). The essential oil of *Zingiber offinalis* Linn (*Zingiberaceae*) as mosquito larvicidal and repellent agent against the filarial vector *Culex quinquefasciatus* Say (Diptera: Culicidae). Parasitol. Res. 102:1289-1291.

Rabha B, Gopalakrishnan R, Baruah I, Singh L (2012). Larvicidal activity of some essential oil hudrolates against dengue and filariasis vectors. E3 J. Med. Res. 1(1):014-016.

Ranaweera SS, Dayananda KR (1996). Mosquito larvicidal activity of Ceyclon citronella (*Cymbopogon nardus* (L.) Rendle) of oil fractions. J. Natn. Sci. Sri Lanka. 24(4):247-252.

Tiwary M, Naik SN, Tewary DK, Mittal PK, Yadav S (2007). Chemical composition and larvicidal activities of the essential oil of *Zanthoxylum armatum* DC (*Rutaceae*) against three mosquito vectors. J. Vect. Borne. Dis. 44:198-204.

World Health Organization (WHO) (2005). Guidelines for labaratory and field testing of mosquito larvicides. Communicable disease control, prevention and eradication, WHO pesticide evaluation scheme. WHO, Geneva; WHO/CDS/WHOPES/GCDPP/1.3.

World Health Organization (WHO) (2012). World Malaria Report.

Zhu L, Tian YJ (2011). Chemical composition and larvicidal effects of essential oil of *Blumea martiniana* against *Anopheles anthropophagus*. Asian. Pac. J. Trop. Med. pp. 371-374.

Zhu L, Tian YJ (2013). Chemical composition and larvicidal activity of essential oil of *Artemisia gilvescens* against *Anopheles anthropophagus*. Parasitol. Res. 112:1137-1142.

Full Length Research Paper

Co-feeding strategy to enhance phytase production in *Pichia pastoris*

Ndayambaje, Jean Bernard and S. Meenakshisundaram*

Centre for Biotechnology, Anna University, Chennai, India.

Process techniques of the methylotrophic yeast *Pichia pastoris* for the production and recovery of heterologous phytase proteins has developed in last 20 years. High expression levels using methanol as induction have been made in the quality of recombinant proteins in the fermenter culture and in the quality of the protein product. This allowed rapidly *P. pastoris* to become the system of choice for the expression of recombinant proteins in yeast. The experimental designs, the methanol/L-alanine co-feeding strategy and optimization of phytase production by *P. pastoris* supported by the optimum levels of variables and lower temperature expression produced high level of phytase activity which could be scaled up to produce phytase for food additives at industrial level. An overall phytase activity was 8632 U/ml, this means 332 fold increase compare to the wild type of phytase. This work demonstrates not only the impact of α-factor prepro secretion signal and efficiency of methanol/L-alanine co-induction strategy for phytase production by recombinant *P. pastoris* Mut[+] strains, but also shows new insights for the expression of bioproduct at lower temperature.

Key words: Phytase, gene expression, *Pichia pastoris*, process optimization, co-feeding strategy, Deglycosylation.

INTRODUCTION

The phosphate (Phytic acid) which is released from non ruminant animals becomes pollutant to the environment and as far as the phytase enzyme for its degradation is needed as food additive (Yu et al., 2012; Xiong et al., 2006; Haefner et al. 2005). For a long time, methanol is used not only as energy and carbon source but also as an inducer of recombinant protein expression (McKinney et al., 2004), and long ago, at high concentrations, it inhibits growth (Zhang et al., 2000). The Glycerol Batch Phase and Glycerol Fed Batch are first fed to the culture to increase biomass concentration and then the culture is switched to methanol to increase productivity, cell density

and also to reduce the induction time. However, the optimal level of protein expression is not achievable with methanol induction alone, due to a partial repression of the *AOX*1 promoter (Faber et al., 2005) and nitrogen source limitation (Callewaert et al., 2001); that is why co-feeding strategy has been applied. The proteolysis of the secreted products and cell death in the high cell density bioreactor cultures is the main limitations when the enzyme is expressed at higher temperature (Li et al., 2006; Pakkanen et al., 2003; Lee et al., 2005; Porro et al., 2005). The expression at lower temperature (20°C) is favorable for efficient heterologous protein expression,

*Corresponding author. E-mail: meenakshi@annauniv.edu. Tel: +91-44-2235 8365, 2254 2298. Fax: +91-44-2235 0299.

and the targeted protein production, specific methanol consumption rate, as well as specific alcohol oxidase (*AOX*1) and its facility in performing many post-translational modifications (Macauley-Patrick et al., 2005).

Therefore, this study focused on lower temperature expression of phytase in *Pichia pastoris* GS115 under co-feeding strategy of methanol/alanine at small scale expression and large scale fermentation. Naturally, the absence of nitrogen is often a limited factor for growing the yeast for enzyme expression at small scale and high cell density fermentation (Heyland, 2010; Fu et al., 2011; Sohn et al., 2010; Sola et al., 2007; Celik et al., 2009). *P. Pastoris* GS115, a Mut^s strains have been found to be useful for production of recombinant protein (Mullaney and Ullah, 2005; Gellissen, 2000; Yang et al., 2004). Through the use of an expression plasmid that contains an α-factor secretory signal sequence, the heterologous proteins are able to be secreted into the medium (Feist et al., 2009). In this work the methanol/alanine induction strategy was applied in 4 L NBS fermenter at 28, 24, and 20°C and the extracellular phytase enzyme was analyzed. The interpretation of the results from methanol/alanine induction strategy was compared with the methanol induction alone together with previously higher temperature expression of phytase by understanding the nitrogen metabolism and the mechanism of energy regeneration. The results and relevant data might be useful for alternative phytase production at industrial level for food additives.

MATERIALS AND METHODS

Bacterial strain and reagents

Aspergillus ficuum NRRL 3135 has been sequenced by TIGR (The Institute for Genomic Research) and was cultivated on Czapek Dox broth and agar, which has the following composition (g/l): sucrose 30, sodium nitrate 2, dipotassium phosphate 1, magnesium sulphate 0.5, potassium chloride 0.5, ferrous sulphate 0.01, agar 15 and the final pH (at 25°C) 7.3±0.2. Restriction enzymes (*EcoRI* and *NotI*) were purchased from NEB. The DNA gel-extraction kit and the plasmid extraction spin mini-prep kit were bought from Qiagen (Germany). Media components such as yeast extract, bacto peptone, bacto tryptone and agar were obtained from Himedia (India). The chemical reagents were obtained from Merck or SRL (India). The *P. pastoris* expression kit was obtained from Invitrogen (SanDiego, CA). Plasmid pPICZαA, which contains *AOX1* promoter, alpha-factor prepro secretion signal and a *Zeocin* selectable resistance marker was used as a yeast. *Escherichia coli* shuttle vector for recombinant phytase expression. *E. coli* Top 10F' was used for construction and propagation of the expression vector. The *P. pastoris* strain GS115 (*his4*) was used for the expression of phytase from the *A. ficuum* strain.

Preparation of mRNA and cDNA construction by reverse transcription polymerase chain reaction (RT-PCR)

Isolation of total RNA from *A. ficuum* NRRL 3135

A. ficuum was grown in a 250 ml baffled flask containing a 50 ml of Czapek Dox broth medium. After 3 days of incubation at 30°C and 200 rpm in shaker flask, the spores of *A. ficuum* were harvested for the extraction of total RNA using the RNeasy Mini Kit from QIAGEN Company with minor modification and the total RNA quantitation assessment was performed with Nano Drop 2000/2000c from (JH BIO Innovations Pvt. Ltd/Thermo scientific, India).

mRNA preparation

The mRNA were performed using T7-Oligo (dT) Promoter Primer; PCR primer to capture the poly (A) tail and 5' SMART IV Oligonucleotide); oligo (dT) linker-primers according to the invitrogen's instructions kit.

cDNA construction

PCR-based gene amplifications were performed using Phusion-HF reaction mix (Finnzymes) and screening PCR reactions were performed using Red-Taq Ready Master mix (BioRad) using the following 20 μl volume reaction; Total RNA (5 μl), oligo (dT) (2 μl), dNTP mix (4 μl), Nuclease free water (4 μl), 5X Buffer RT (2 μl), DTT (1 μl), RNase free water (1 μl), Supercript II RNase-H.RT (1 μl). The phytase gene was amplified using the following primers: The upstream primer 5'CCG GAA TTC CTG GCA GTC CCC GCC TCG AGA 3' with *EcoRI* restriction enzyme site and the downstream primer 5' TAA AGC GGC CGC CTA TGC AAA ACA CTC CGC 3' with *NotI* restriction enzyme site. The ligation mixture was transformed into competent *E. coli* DH5α. The transformants were grown in Luria Bertani broth with low salt concentrations (LB-LS) plates then supplemented with *Zeocin* (1 μg/mL). The clone that contained the PCR product was verified by restriction enzyme digestion, agarose gel electrophoresis, and sequencing.

Shake flask expression study of phytase

Shake flask expression of the transformed *P. pastoris* GS115 clones were inoculated in 250 ml conical flasks with the YPG complex media in duplicates. The shake flasks containing 50 ml of the media were inoculated with 1 ml of transformed *P. pastoris* culture, and grown into 3 mL YPD medium and incubated first at 28°C, and then decreased to 20°C. The culture was transferred into 50 mL YPG medium under *Zeocin* resistance. Once the OD$_{600}$ reached between 20 and 25; the cells were pelleted and suspended in BMMY (YNB-700 μl, Biotin-14 μl, 100 mM potassium phosphate pH 6.0 to 700 μl and distilled water - 7 ml) and later in unbuffered MMH medium. Initially, uninduced sample was taken and induction was carried out with methanol/L-alanine (0.5% for the first day, 1% for subsequent day; 0.3925, 0.7185 g/l/h, respectively for subsequent days till 7 days). The samples were taken and centrifuged 10,000 X *g* for 10 min at 4°C and the supernatant was taken for further analysis. From these clones, highly expressed clone was chosen for reactor studies.

Fermentation strategy

Transformed *P. pastoris* GS115 was first cultured in a 500 ml shaker flask containing 100 ml BSM at 28°C until an OD$_{600}$ value of around 35 had been reached. For high cell-density fermentation, *P. pastoris* GS115, the seed culture equal to 400 ml was added into a NBS fermentor (NBS BioFlo 415 Benchtop SIP fermentor) containing sterilized 4 L of BSM. The components of 1 L BSM medium: $CaSO_4 \cdot 2H_2O$ - 0.46 g/l; K_2SO4 - 9,1 g/l; $MgSO_4 \cdot 7H_2O$ - 7.45 g/l; KOH - 2.06 g/l; Glycerol - 40 g/l; H_3PO_4 - 26.7 ml/l; Histidine, 0.4 g/l. After sterilization and cooling down to 28°C for getting enough biomass, temperature was kept to 20°C throughout the cultivation

time and pH of the medium was adjusted to 4.8 with 25% ammonium hydroxide and 6 ml PTM$_1$ trace salts/liter of basal salts medium was added aseptically. Composition of PTM$_1$ solution: CuSO$_4$ - 6 g/l; NaI - 0.08 g/l; MnSO$_4$ - 3 g/l; NaMo - 0.2 g/l; H$_3$BO$_3$ - 0.02 g/l; CoCl$_2$ - 0.5 g/l; ZnCl - 20 g/l; FeSO$_4$ - 65 g/l; H$_2$SO$_4$ - 5 ml/l; Biotin - 0.2 g/l (Blumhoff et al., 2013). The stirring speed, airflow and the pH were monitored according to the following conditions: Temperature (28 to 20°C), Dissolved oxygen (>20%), Methanol for carbon source and L-alanine as nitrogen source, pH (4.8) was adjusted using 25% NH$_4$OH and 88% H$_3$PO$_4$, agitation (200 to 800 rpm), aeration (0.1 to 1.0 vvm for glass fermenters), antifoam (the minimum PPG is needed to eliminate foam). Another 40 g/l glycerol was fed to the reactor to increase the biomass before the induction phase. Cells were collected by centrifugation and resuspended in 1.5 ml tubes for further analysis. Feeding medium for induction: Pure methanol (0.5 to 8 ml/l/h, PTM1:4.5 ml/l at rate of 2 ml/L/day, L-alanine co-feeding rates: 0.3925 and 0.7185 g/L/h. Samples were taken periodically throughout the fermentation time for phytase assay and protein analyses.

Phytase assay

Phytase activity was determined according to the report of Bae et al. (1999) with minor modification. Briefly, 75 µl of enzyme solution was incubated with 300 µl substrate solution (1.5 mM sodium phytate in 0.1 M sodium acetate buffer, pH 5.0) at 37°C for 20 min. The reaction was stopped by adding a volume of 375 µl of 5% (w/v) trichloroacetic acid. The released inorganic phosphate was analyzed by adding 375 µl of a coloring reagent (freshly prepared by mixing four volumes of 1.5% (w/v) ammonium molybdate in a 5.5% (v/v) sulfuric acid solution and one volume of a 2.7% (w/v) ferrous sulfate solution) and the solution's absorbance at 595 nm was measured using a Versamax microplate reader. The activity of the strain was analyzed where "one phytase unit was defined as the activity that releases 1 µmol of inorganic phosphorus (*Pi*) from sodium phytate per minute at 37°C". One strain was selected because of its high activity and SDS-PAGE was used to check the size of the phytase protein.

Digestion of glycopeptides with PNGase F

The peptide-N4-(N-acetyl-β-glucosaminyl)-asparagine amidase F cleaves selectively N-glycans from the asparagine rest of the peptide (Hanson et al., 2009). The reaction was started by denaturing glycoprotein buffer with 0.5% SDS, 40 mM DTT at 100°C for 10 min. After an addition of NP-40 and G7 reaction buffer, two-fold dilutions of *PNGase F* were added and the reaction mix was incubited for 1 h at 37°C and the separation of reaction products were visualized by SDS-PAGE.

RESULTS

Amplification and cDNA construction by RT-PCR

The amplification of 1.3 kb phytase gene was resolved by 1% agarose gel. The *EcoR*I and *Not*I both cut the double-stranded DNA at specific recognition nucleotide sequences. Ligation PCR was performed with T4 DNA ligase, *AOX*1 forward and phytase reverse primers were used for the confirmation PCR reaction. Phytase gene 1.3 kb has been sent to synergy lab for sequencing and it was compared with previously isolated phytase genes.

This confirmed that our recombinant phytase has been successfully inserted into chromosome of *P. pastoris* GS115.

Small scale expression and high cell density fermentation

The phytase activity in shaker flask using BMMY medium under methanol/L-alanine induction was 298 U/ml and it increased gradually up to 148 h cultivation time. The inoculum for high cell density fermentation was prepared from glycerol stocks maintained at -20°C. The frozen cells were thawed and inoculated in 3 ml test tube YPG medium. After 24 h, 1 ml culture was inoculated into 20 ml YPG medium in a sterile 100 ml flask and grown for 48 h. Both test tube and flask were kept in incubator at 28°C and 200 rpm. The inoculum equals to 20 ml was transferred aseptically to 200 ml sterile Basal salts medium in 500 ml Erlenmeyer flasks. The culture was grown for 36 h at 28°C and 200 rpm until the inoculum OD$_{600}$ reached approximately 30. The fermenter was inoculated with 200 ml culture from shake flasks and the biomass concentrations were monitored with time. As the culture grew, the DO decreased and when the glycerol in the medium was completely consumed the DO rose sharply. This indicated the end of the initial batch phase and the second step of glycerol 98% fed-batch was started; after 45 h, it reached OD$_{600}$ = 217. The optical density growth curve and the activity were found to be proportional during the induction time. The methanol/L-alanine co-feeding strategy increased the phytase activity to 8632 U/ml (Figure 1) compare to 2711 U/ml activity of phytase obtained under cells grown on methanol induction alone.

The productivity increased to 233 (U/ml/h) and it was obtained from the ratio of the activity and the induction time hours and the specific productivity arose to a level of 1058 (U/g/h) and it was calculated from a thousand productivity to biomass (Figure 2). The effect of L-alanine as nitrogen source and methanol as carbon source increased biomass to 53 g/l. The biomass decreased compared to that obtained at higher temperature expression and it affected positively the phytase concentration because it increased remarkably to 13.6 mg/ml after 144 h cultivation time (Figure 3).

SDS-PAGE analysis

The cells were pelleted and suspended in BMMY (YNB-700 µl, Biotin-14 µl, 100 mM potassium phosphate, pH 6.0 to 700 µl and distilled water – 7 ml) medium. The phytase was induced by methanol/L-Alanine (0.5% for first day and 1%, 0.3925 and 0.7185 g/L/h, respectively for subsequent days till 7 days). Due to the heavy glycolsylation, the expressed phytase was found to have molecular sizes of around 120, 116, and 66 kDa and this showed that phytase is a highly glycosylated proteins
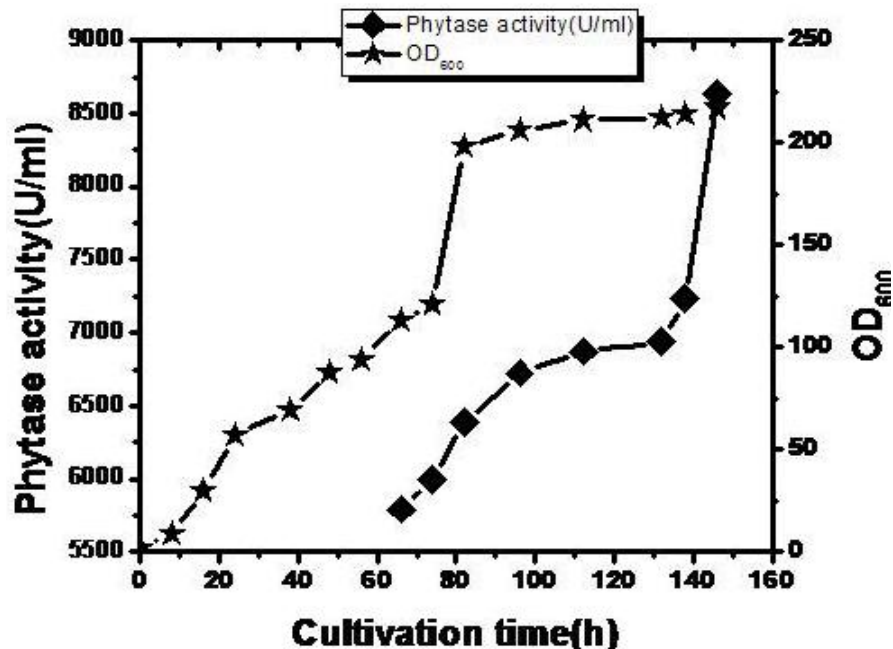
Figure 1. Phytase activity and OD_{600nm} during cultivation time on methanol/L-alanine co-feeding strategy. The Glycerol Batch Phase and Glycerol Fed Batch last 48 h. After glycerol phases, feeding was switched to co-feeding methanol/L-alanine, which induced phytase expression, then the activity and expression level of phytase were increased up to 146 h cultivation time.
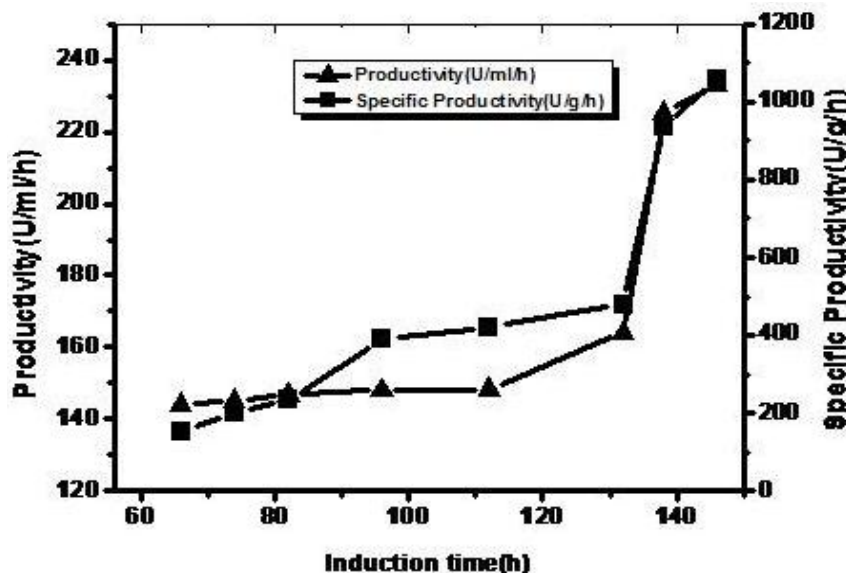
Figure 2. Productivity (U/ml/h) and Specific Productivity (U/g/h) during induction time of phytase production under methanol/L-alanine at 20°C.

(Figure 4).

DISCUSSION

Isolation of phytase gene (*phyA*) from *A. ficuum* NRRL 313, construction of recombinant *pPICZαA*-Phytase, inte-

gration and expression of phytase into *P. pastoris* GS115 was successful. Different reactors were run in different conditions at lower temperature expression of phytase (20°C) some reactors with nitrogen and others without nitrogen sources and this confirmed that the nitrogen source was one of the limiting factors for growth in all the
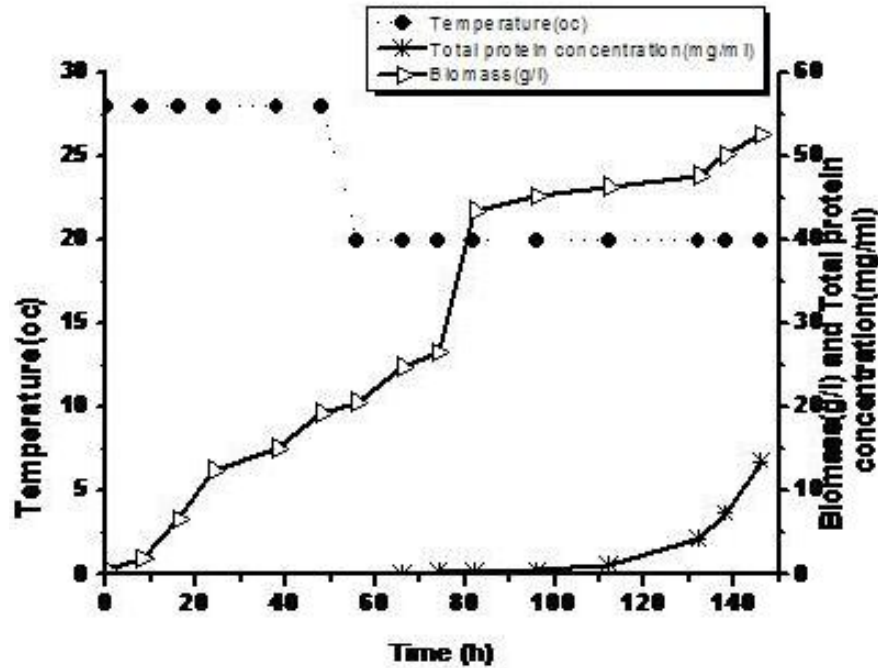
Figure 3. Biomass (g/l), Temperature ($^{\circ}$C) for the time of Glycerol Batch Phase and Glycerol Fed Batch; Phytase concentration (mg/ml) under methanol/L-alanine co-feeding induction strategy.
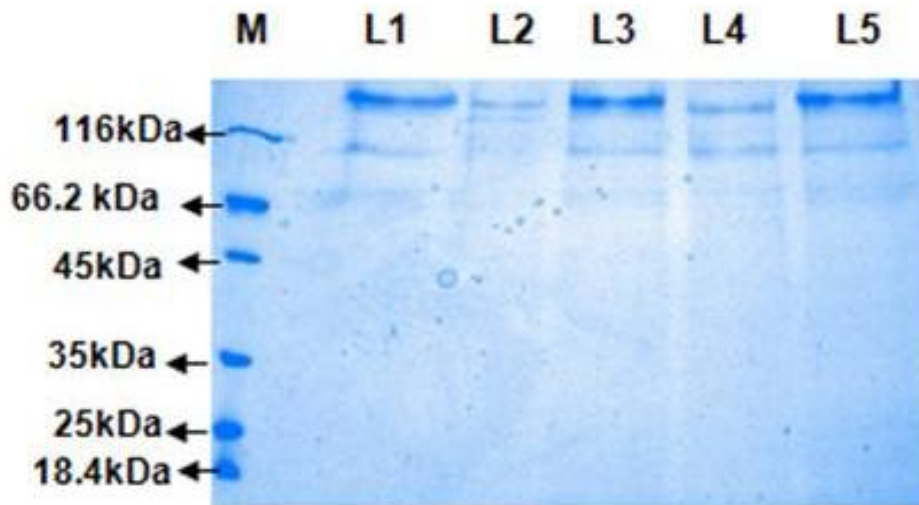
Figure 4. SDS-PAGE analysis of recombinant phytase expressed in BMMY medium. M: Unstained protein Ladder Markers. (L1, L2, L3, L4, L5): Phytase protein before deglycosylation showing multiple bands. **Note:** The molecular weight of phytase protein after deglycosylation with *PNGase* F was approximately 45 kDa as determined by SDS-PAGE (Figure 5).

continuous cultures. Cells grown under methanol/L-alanine limitations had high activity compared to cells grown on methanol feeding alone. The mechanism of methanol/L-alanine co-feeding strategy increased the enzyme activity to 8632 U/ml under the control of *AOX1* promoter and

α-factor signal peptide meaning three times com-pared to methanol induction alone in the fermenter and 29 times from the shake flask expression. The alternative reason for producing higher yield under methanol/L-alanine co-feeding strategy might happen around pyruvate, acetyl co-
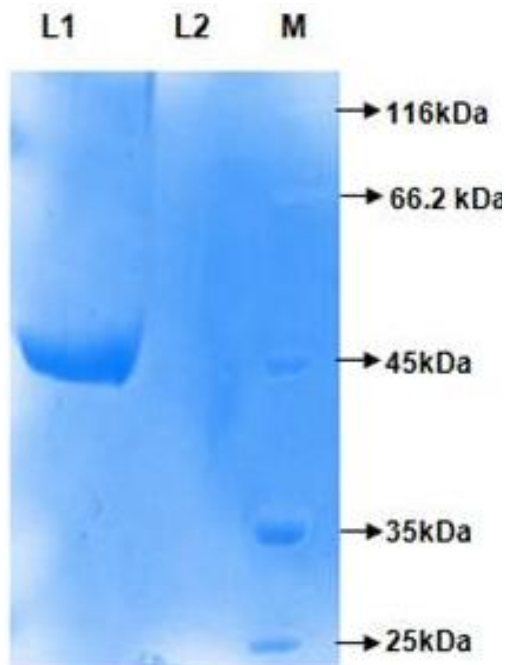
Figure 5. L1: Optimized Phytase protein after deglycosylation with PNGaseF (New England Biolabs). L2: Negative control with GS115-*pPICZαA*. M: Unstained protein ladder markers.

enzyme A, glyoxylate, and α-ketoglutarate via increased levels of *ALT1, DAL7, PYC1, GDH2,* and *ADH5* and decreased levels of *GDH3*, *CIT2*, and *ACS1* transcripts (Usaite et al., 2006) and also the physiological impacts of L-alanine to the cell is to supply enough nitrogen source (Fu et al., 2013).

The novel phytase that has a high extracellular activity and other characterizations (for example, temperature stability, wide pH optima, etc.), which are necessary for the commercial utilization using the methylotrophic yeast, *P. pastoris* is now available to be scaled up to the industrial level for food additives. The wide range of promoters available, as well as selectable markers, secretion signals, methods for coping with proteases and a better understanding of glycosylation patterns, are powerful to *P. pastoris* as a high efficient expression system available. The well-defined process protocols and some degree of process optimization are required to achieve maximum production of the heterologous protein. In fact, yield and activity are often dependent upon the parameters of the culture vessel (pH, temperature and O_2 availability), and they are also dependent on the residual L-Alanine and methanol concentrations. These factors can be closely monitored to ensure the exact conditions required. *P. pastoris* is able to add both O-linked and N-linked carbohydrate moieties to secreted proteins (Wang et al., 2005); the reason why phytase presented different bands size but after deglycosylation, the exact protein size of phyatse (*phyA*) became 45 kDa.

However, the induction at low temperature minimized extracellular proteolysis (Heyland, 2010) though it led to a high operation cost but the activity and protein concentration increased; and the toxicity of methanol was reduced compared to the induction at higher temperature and methanol feeding alone because there was too much methanol consumption. The cost of methanol and its hazardous substance is very high due to its high flammability and toxicity (Zou et al., 2006); additionally, cells growing on methanol have a very high oxygen consumption, which usually requires the addition of pure oxygen to the culture, increasing the cost of the process and limiting the cultivation capacity at high scale. Methanol was a cheaper and readily available substrate at the time the *P. pastoris* system was developed (Jahic et al., 2002). However, the wide range of applications in the field of protein production developed over the years have revealed the need for a more controllable, less volatile, and less flammable substrate for induction.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

Bae HD, Yanke LJ, Cheng KJ, Selinger LB (1999). A novel staining method for detecting phytase activity. J. Microbiol. Methods 39(1): 17-22.

Blumhoff M, Steiger MG, Marx H, Mattanovich D, Sauer M (2013). Six novel constitutive promoters for metabolic engineering of *Aspergillus niger*. Appl. Microbiol. Biotechnol. 97(1): 259-267.

Callewaert N, Laroy W, Cadirgi H (2001). Use of HDEL-tagged *Trichoderma reesei* mannosyl oligosaccharide 1, 2-α-D-mannosidase for N-glycan engineering in *Pichia pastoris*. FEBS Lett. 503(2): 173-178.

Celik E, Calik P, Oliver SG (2009). Fed-batch methanol feeding strategy for recombinant protein production by *Pichia pastoris* in the presence of co-substrate sorbitol. Yeast 26:473-484.

Faber KN, Harder W, Ab G, Veenhuis M (2005). Methylotrophic yeasts as factories for the production of foreign proteins. Yeast 11(14): 1331-1344.

Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009). Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol. 7:129-143.

Fu XY, Zhao W, Xiong AS, Tian YS, Peng RH (2011). High expression of recombinant *Streptomyces* sp. S38 xylanase in *Pichia pastoris* by codon optimization and analysis of its biochemical properties. Mol. Biol. Rep. 38: 4991–4997.

Gellissen G (2000). Heterologous protein production in methylotrophic yeasts. Appl. Microbiol. Biotechnol. 54(6):741-750.

Haefner, S., A. Knietsch, E. Scholten, J. Braun, M. Lohscheidt, and O. Zelder (2005). Biotechnological production and applications of phytases. Appl. Microbiol. Biotechnol. 68:588-597.

Hanson SR, Culyba EK, Hsu TL, Wong CH, Kelly JW, Powers ET (2009). The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. Proc. Natl. Acad. Sci. USA 106(9): 3131-3136.

Heyland J (2010). Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. Biotechnol. Bioeng. 107(2): 357-368.

Jahic M, Rotticci-Mulder M, Martinelle M, Hult K, Enfors SO (2002). Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein. Bioprocess Biosyst. Eng. 24: 385-393.

Lee CCW, Tina G, Wong DWS (2005). An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*. Plasmid 54(1): 80-85.

Li H, Sethuraman N, Stadheim TA, Zha D, Prinz B, Ballew N, Bobrowicz P, Choi BK, Cook WJ, Cukan M, Houston-Cummings NR, Davidson R, Gong B, Hamilton SR, Hoopes JP, Jiang Y, Kim N, Mansfield R, Nett JH, Rios S, Strawbridge R, Wildt S, Gerngross TU (2006). Optimization of humanized IgGs in glycoengineered *Pichia pastoris*. Nat. Biotechnol. 24(2): 210-215.

Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005). Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22(4): 249-270.

McKinney J, Knappskog PM, Pereira J (2004). Expression and purification of human tryptophan hydroxylase from *Escherichia coli* and *Pichia pastoris*. Protein Expr. Purif. 33(2): 185-194.

Mullaney EJ, Ullah AH (2005). Conservation of cysteine residues in fungal histine acid phytases. Biochem. Biophys. Res. Commun. 328: 312-179.

Pakkanen O, Haemaelaeinen ER, Kivirikko KI, Myllyharju J (2003). Assembly of stable human type I and III collagen molecules from hydroxylated recombinant chains in the yeast *Pichia pastoris* Effect of an engineered C-terminal oligomerization domain foldon. J. Biol. Chem. 278(34): 32478-32483.

Porro D, Sauer M, Branduardi P, Mattanovich D (2005). Recombinant protein production in yeasts. Mol. Biotechnol. 31(3): 245-259.

Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY (2010). Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. Biotechnol J. 5(7):705-15.

Sola A, Jouhten P, Maaheimo H, Sanchez-Ferrando F, Szyperski T, Ferrer P (2007). Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. Microbiology 153:281-290.

Usaite R, Patil KR, Grotkjær T, Nielsen J, Regenberg B (2006). Global transcriptional and physiological responses of *Saccharomyces cerevisiae* to ammonium, L-alanine, or L-glutamine limitation. Appl. Environ. Microbiol. 72(9): 6194-6203.

Wang HN, Wu Q, Zhao HX, Chen H, Liu P (2005). Secretory expression of *Bacillus subtilis* phytase phyC in *Pichia pastoris*. J. Zhejiang Univ. (Agric. Life Sci.) 31:621-627.

Xiong AS, Yao QH, Peng RH, Zhang Z, Xu F, Liu JG, Han PL, Chen JM (2006). High level expression of a synthetic gene encoding *Peniophora lycii* phytase in methylotrophic yeast *Pichia pastoris*. Appl. Microbiol. Biotechnol.72: 1039-1047.

Yang G, Zhou H, Lu Y, Lin Y, Zhou S (2004). Comparing expression of different forms of human DNA topoisomerase I in *Pichia pastoris*. Enzyme Microb. Technol. 34(2): 139-146.

Yu S, Cowieson AJ, Gilbert C, Plumstead P, Dalsgaard S (2012). Interactions of phytate and myo-inositol phosphate esters (IP1-5) including IP5 isomers with dietary protein and iron and inhibition of pepsin. J. Anim. Sci. 90: 1824-1832.

Zhang W, Inan M, Meagher MM (2000). Fermentation strategies for recombinant protein expression in the methylotrophic yeast *Pichia pastoris*. Biotechnol. Bioproc. Eng. 5: 275–287.

Zou LK, Wang HN, Pan X, Xie T, Wu Q, Xie ZW, Zhou WR (2006). Design and expression of synthetic phyC gene encoding the neutral phytase in *Pichia pastoris*. Acta Biochim. Biophys. Sin. (Shanghai) 38:803-811.

Review

Current trends in genetic manipulations to enhance abiotic and biotic stresses in tobacco

Asmamaw Tesfaw* and Tileye Feyissa

Addis Ababa University (AAU) Arat Kilo P.O. Box 1176 Addis Ababa, Ethiopia.

Environmental stresses, both abiotic and biotic, are the main factors reducing crops productivity. Biotic stresses caused by fungi, viruses, bacteria and insects are the major threats. However, abiotic stresses have more adverse effects on crop yield and crop survival. Hitherto, tolerant plants were mainly produced by classical breeding techniques. Success in breeding for better adapted varieties to abiotic and biotic stresses depends on the concerted efforts of various research domains including plant and cell physiology, molecular biology, genetics and breeding. However, such process is time consuming. The production of transgenic plants by genetic engineering allows a much faster progress. This review focuses on stress tolerance on tobacco. Tolerance to salt, drought, water stress, low and high temperature, heavy metals, pathogens and pests were enhanced via overexpression of specific stress related genes from different sources.

Key words: Overexpression, genetic engineering, abiotic factors, biotic factors.

INTRODUCTION

Out of a US$1.3 trillion annual food production capacity worldwide, the biotic stresses (insects, diseases and weeds) cause 31 to 42% loss (US$500 billion), with an additional 6 to 20% (US$120 billion) post harvest loss due to insects, fungal rots and bacterial rots (FAO, 2005). FAO (2005) also investigated 6 to 20% (US$120 billion) loss by abiotic stresses (drought, flood, frosts, nutrient deficiencies, various soil and air toxicities). Goel and Madan (2014) reported that abiotic stresses lead to more than 50% of yield reduction. These huge losses exert high pressure in world food security program.

Plants are sessile organisms which are constantly exposed to a variety of biotic and abiotic stresses in their external environment. Abiotic stress causes cellular dehydration and accumulation of reactive oxygen species (ROS) such as hydrogen peroxide, which may function as signal to counteract these stresses, for instance by increasing endogenous osmotic pressure to prevent further water loss from the cell (Doltchinkova et al., 2013). In general, ROS are one of the major damaging factors that arise when plants are exposed to various abiotic stresses, including nutrient excess/depletion, flooding drought, desiccation, extreme temperatures, light/dark cycles, high salinity, ozone, anoxia and herbicides (Choet et al., 2013). Accordingly, an unfortunate consequence of salinity stress in plants is the excessive generation of ROS intermediates, such as superoxide radicals (O_2^-), hydrogen peroxide (H_2O_2) and

*Corresponding author. E-mail: astesfa@yahoo.com. Tel: +251- 911 956817 or +256 75 855 1340.

Table 1. Gene manipulation to enhance abiotic stress tolerance in transgenic tobacco.

Gene/its product	Source	Role	Reference
GmERF7	Soybean	Enhanced salt tolerance	Zhai et al. (2013)
MDHAR	Acerola	Antioxidative	Eltelib et al. (2012)
ALDRXV4	*Xerophyta viscose*	Better salinity stress tolerance	Kumar et al. (2013)
AtNHX1	*Arabidopsis*	Increased salt tolerance	Zhou et al. (2011)
DgNAC1	chrysanthemum	Improved salt tolerance	Liu et al. (2011)
AVP1	*Arabidopsis*	Tolerated to shortage of water	Arif et al. (2013)
ALDRXV4	*Xerophyta viscosa*	Survived longer period of water deficiency	Kumar et al. (2013)
TaEXPB23	Wheat	Losing water more slowly	Li et al. (2011)
PtADC	*Poncirus trifoliata*	Reduce ROS	Wang et al. (2011)
LOS5/ABA3	*Arabidopsis*	Maintaining high water content	Yue et al. (2011)
TaEXPB23	*Arabidopsis*	Improved the water-stress tolerance	Li et al. (2013)
BcBCP1	*Boea crassifolia*	Tolerance to osmotic stress	Wu et al. (2011)
CbCOR15b	Shepherd's purse	Cold tolerance	Wu et al. (2012)
GbCBF1	Cotton	Enhanced chilling tolerance	Hui-Ming et al. (2011)
ZmMPK4	Maize	Low temperature tolerance	Zhou et al. (2012)
CfCBF3	Sweet pepper	Low temperature tolerance	Yang et al. (2011)
ZFP177	Rice	High temperature tolerance	Grover et al. (2013)
WRKY40	Pepper	High temperature tolerance	Dang et al. (2013)
MxCS1	*Malus xiaojinensis*	Improved Fe stress	Han et al. (2013)
TaVP1	Wheat	Enhanced Cd tolerance	Khoudi et al. (2012)

hydroxyl radicals (OH^-). ROS overproduction affects membrane potential and other essential macromolecules, such as photosynthetic pigments, carbohydrates, proteins, DNA and lipids inhibiting photosynthetic capacity.

Plants developed an array of complex mechanisms to rapidly sense environmental changes. Plants also developed a variety of mechanisms to adapt themselves to ever changing environments and these mechanisms are mediated through multiple signal transduction pathways acting in a global signal network. Typically, mitogen-activated protein kinase (MAPK) cascades are universal signaling pathways employed in such responses (Zhou et al., 2012). Plants adapt to unfavorable environmental conditions by accumulating highly soluble organic compounds of low molecular weight like sugars, sugar alcohols, amino acids and amino acid derivatives (Doltchinkova et al., 2013). In addition, plants detoxify ROS through a combination of antioxidants (ascorbate (AsA) and glutathione (GSH), and antioxidative enzymes [superoxide dismutase (SOD); ascorbate peroxidase (APX); monodehydroascorbate reductase (MDHAR); dehydroascorbate reductase, (DHAR); and catalase (CAT)] (Choe et al., 2013; Peshev et al., 2013). Antioxidative enzymes are involved in chloroplasts, cytosol and mitochondria that are known sources of ROS generation. In addition, transgenic plants with increased contents of compatible solutes show a strong potential for improving their stress tolerance (Doltchinkova et al., 2013). This review summarizes the gene sources and the effects on stress tolerance after transforming these genes into tobacco.

ABIOTIC FACTORS

Abiotic factors are yield limiting factors that potentially reduce crop productivity. Abiotic factors include salinity, drought, osmotic stress, heavy metals and natural and artificial toxic chemicals. Before the era of modern biotechnology, these problems were overcome by exposing plants for long periods to allow the plants to adapt, finally creating tolerant organisms. Plant biotechnology allows a quick introduction of a tolerance-related gene from a certain organism into susceptible plants (Table 1). Such genes are typically overexpressed by using promoters to enhance their functions. For instance, transgenic tobacco plants overexpressing *OsCBSX4* [a gene that produce a component of cystathionine β-synthase domain containing proteins (CDCPs)], obtained from rice (*Oryza sativa* L.) exhibited improved tolerance against salinity, heavy metals and oxidative stress (Singh et al., 2012). The higher accumulation of OsCBSX4 protein in *OsCBSX4* overex-pressing transgenic plants and the exhibition of higher abiotic stress tolerance than wild type (WT) plants suggest a role in abiotic stress tolerance in plants.

Salt tolerance

Salinity deteriorates agricultural land and reduces crop

yield. The land affected by salinity is estimated to be about one-third of the world's cultivated land (Eltelib et al., 2012). Accordingly, plants have evolved a variety of mechanisms to deal with high salinity. For example, a low Na^+ content in the cytoplasm is maintained by compartmentation of Na^+ into the vacuole (Zhou et al., 2011). Vacuolar Na^+ compartmentation has two basic roles in salt tolerance physiology. First, it keeps Na^+ away from the sites of metabolism, and second, it increases the osmolarity of the cell to regulate cytoplasmic pH. These are attained using Na^+/H^+ antiporter that pumps excess vacuolar Na^+ into vacuoles, and membrane potential is created due to a charge difference. The membrane potential and H^+-inorganic pyrophosphatase provides energy for Na^+/H^+ antiporter (Zhou et al., 2011) and the energy establishes electrochemical H^+ gradient between the cytoplasm and the vacuole.

Ethylene-response factors (ERFs) play an important role in regulating gene expression in plant responses to biotic and abiotic stresses. Zhai et al. (2013) isolated a new ERF transcription factor, *GmERF7*, from soybean. Overexpression of *GmERF7* in tobacco plants led to higher levels of chlorophyll and soluble carbohydrates and a lower level of malondialdehyde compared to wild-type tobacco plants under salt stress conditions, indicating that *GmERF7* enhanced salt tolerance in transgenic plants.

Eltelib et al. (2012) introduced a monodehydro-ascorbate reductase (MDHAR) cDNA from acerola (*Malpighia glabra*) into tobacco plants via an *Agrobacterium*-mediated gene delivery system. Trans-genic plants accumulated higher level of ascorbate (AsA) and showed higher MDHAR activities as compared with the control plants. These effects were also associated with a greater tolerance to salt stress in transgenic tobacco plants. Transgenic tobacco plants over-expressing aldose reductase (*ALDRXV4*) cDNA, obtained from *Xerophyta viscose*, showed better salinity stress tolerance than WT plants (Kumar et al., 2013).

Zhou et al. (2011) introduced *AtNHX1* (a vacuolar Na^+/H^+ antiporter gene) from *Arabidopsis thaliana* in tobacco. Salt tolerance was better in the *AtNHX1* over expression plants than WT plants, with higher germination rates and successful seedling establishment in the presence of toxic concentrations of NaCl. The authors also proposed that simultaneous overexpression of V-ATPase and vacuolar Na^+/H^+ antiporter might be appropriate to produce plants with a higher salt tolerance ability.

Furthermore, overexpression of a novel Chrysanthemum (*Dendronthema grandiform*) NAC transcription factor gene in tobacco was studied by Liu et al. (2011). A full-length cDNA designated *DgNAC1* has been isolated from Chrysanthemum and it was overexpressed in tobacco to assess its ability to improve stress tolerance. Indeed, it was found that 35S:DgNAC1 transgenic tobacco plants exhibited a markedly increased tolerance to salt stress.

Drought tolerance

Desiccation tolerance has been defined as the ability of an organism to equilibrate its internal water potential with that of moderately dry air, and then resume normal function after rehydration (Arif et al., 2013). Desiccation tolerant flowering plants require a slow drying time to activate mechanisms that protect membranes and organelles during desiccation. Thick cell walls, few vacuoles and deep and closed stomata were reported in tobacco plants overexpressing *Arabidopsis* vacuolar pyrophosphatase (AVP1), whereas WT tobacco showed larger vacuoles and kept their stomata open (Arif et al., 2013). As a result, the transgenic tobacco tolerated water shortage (dessication). In addition, AVP1 transformed tobacco showed better growth (increase in size and weight of shoots and capsules) and they produced more seeds than WT plants. Overexpression of a levansucrase gene from *Bacillus subtilis* in tobacco led to increased drought tolerance (Pilon-Smits et al., 1995), and similar stress tolerance effects were obtained by using other fructan synthesizing genes in a diverse array of plant species (Keunen et al., 2013).

In another study, Kumar et al. (2013) reported the efficacy of an aldose reductase (ALDRXV4) enzyme from *Xerophyta viscosa* Baker in enhancing the prospects of tobacco's survival under abiotic stress. Transgenic tobacco plants overexpressing *ALDRXV4* cDNA showed an alleviation of NaCl and mannitol induced abiotic stress. The transgenic plants survived longer periods of water deficiency than the WT plants. The increased synthesis of aldose reductase in transgenic plants correlated with reduced methylglyoxal and malondialdehyde accumulation and an elevated level of sorbitol under stress conditions. In addition, the transgenic lines showed better photosynthetic efficiency, less electrolyte damage, greater water retention, higher proline accumulation and a more favorable ionic balance under stress conditions.

Plants growth is the result of cell division and cell enlargement. Plants growth is regulated by cell extensions. Drought stress restrains plant growth by retarding both cell division and extension (Gao et al., 2007). Expansins are proteins that are the key regulators of wall extension during plant growth and they mediate pH dependent wall loosening, probably by disrupting hydrogen bonds between cellulose and matrix glycans (Li et al., 2011). *TaEXPB23* is a wheat expansin gene that was transformed into tobacco. The results indicate that the transgenic tobacco lines lost water more slowly than the wild-type plants under drought stress; their cells could sustain a more integrated structure under water stress than that of WT plants (Li et al., 2011).

In a study of Wang et al. (2011), *PtADC* (an arginine decarboxylase gene), isolated from *Poncirus trifoliata*, was introduced into tobacco to investigate its function in drought tolerance. It was demonstrated that the transgenic plants showed an improvement in dehydration

and drought tolerance. Under dehydration stress conditions, ROS accumulation was much lower in the transgenic tobacco plants as compared with WT plants. Similarly, Yue et al. (2011) transformed *Arabidopsis LOS5/ABA3* into tobacco. They found that transgenic plants showed less wilting, maintained higher water content and showed better cellular membrane integrity as compared with control plants. Moreover, the transgenic plants accumulated higher quantities of ABA and proline, and they exhibited higher activities of antioxidant enzymes.

Water stress tolerances

Water stress affects plant growth and significantly decreases crop productivity. Plants respond to water stress with physiological and developmental changes. Therefore, studying the molecular and physiological mechanisms by which plants adapt to water stress is important for improving agricultural practices.

As stated before, expansins are the key regulators of cell wall extension during plant growth. Li et al. (2011) produced transgenic tobacco plants with increased tolerance to water stress by overexpressing the wheat expansin gene *TaEXPB23* driven by the constitutive 35S cauliflower mosaic virus (CaMV) promoter. However, the growth and development of *35S::TaEXPB23* transgenic tobacco plants were altered under normal growth conditions, with a faster growth rate at the seedling stage, earlier flowering and maturation, and a shorter plant height compared to WT plants. Next, Li et al. (2013) altered cellular characteristics and carbohydrate metabolism in *35S::TaEXPB23* transgenic tobacco plants. During carbohydrate metabolism, lower starch accumulation was obtained due to higher amylase activity in the leaves of *35S::TaEXPB23* compared with WT plants. They also generated transgenic *Arabidopsis* plants that showed the same phenotype as the transgenic tobacco plants, which may have resulted from the altered expression of several flowering-related genes; after which, they produced TaEXPB23 transgenic tobacco plants using the stress-inducible *RD29A* promoter. The use of this promoter reduced the negative effects of *TaEXPB23* on plant growth and development. The *RD29A::TaEXPB23* transgenic tobacco plants had greater tolerance to water stress than WT plants, as determined by examining physiological and biochemical parameters. Therefore, the use of stress-inducible promoters, such as *RD29A*, may minimize the negative effects of constitutive transgene expression and improve the water-stress tolerance of plants.

Overexpression of *BcBCP1*, a phytocyanin-related early nodulin-like gene from *Boea crassifolia*, in tobacco under the control of CaMV 35S promoter enhanced tolerance to osmotic stress, as indicated by the less impaired growth, less damaged membrane integrity and lower lipid peroxidation levels after osmotic stress (Wu et al., 2011). Transgenic tobacco lines overexpressing *BcBCP1* showed higher photosynthetic rates, higher antioxidant enzyme activities and higher cytosyl ascorbic peroxidase transcription levels than non-transgenic tobacco plants, both under normal conditions and under osmotic stress.

Low temperature tolerance

Low temperature is one of the most important abiotic factors limiting growth and productivity of crop plants. Cold regulated (COR) genes are low temperature-responsive genes regulated by the C-repeat binding-factor (CBF) signaling pathway (Hui-Ming et al., 2011; Yang et al., 2011; Wu et al., 2012). When *CbCOR15b* from shepherd's purse (*Capsella bursa-pastoris*), was expressed in transgenic tobacco plants under chilling and freezing temperatures, transformants were found more tolerant under cold (Wu et al., 2012), judged on electrolyte leakage experiments, relative water content, glucose content and phenotypic observations. Similarly, a CBF gene named *GbCBF1* was isolated from cotton and introduced into tobacco (Hui-Ming et al., 2011). They found that under low temperature stress, the electrolytic leakage rate of transgenic tobacco was lower than that of the WT tobacco. However, free proline and soluble sugar contents of transgenic tobacco were higher than those of the WT tobacco. These results indicate that *GbCBF1* enhances cold tolerance in transgenic tobacco.

Various studies confirmed that the CBF family of transcription factors plays key roles in regulating cold stress responses. Yang et al. (2011) isolated CBF3 from sweet pepper (*Capsicum frutescens*) and transformed it into tobacco. Overexpression of *CfCBF3* under the control of the CaMV35S promoter in tobacco induced expression of orthologs of CBF3-targeted genes and increased chilling tolerance without a dwarf phenotype. Multiple biochemical and physiological changes, such as increased levels of proline and soluble sugars and lower contents of ROS, were observed in transgenic plants associated under chilling condition. Sucrose specific signaling pathways may be involved in such stress responses (Van den Ende and El-Esawe, 2014). In addition, overexpression of *CfCBF3* resulted in higher level of total unsaturated fatty acids, especially in phosphatidylglycerol (PG). During exposure to chilling stress, the transgenic lines were less susceptible to chilling-induced photoinhibition than WT plants. These results suggest that overexpression of *CfCBF3* led to modification of the fatty acid unsaturation and alleviated the injuries under chilling stress.

Another universal signaling pathway involved in responses to low temperature is the mitogen-activated protein kinase (MAPK) cascade. *ZmMPK4* is a mitogen-activated protein kinase gene from maize. Transgenic

tobacco overexpressing *ZmMPK4*, accumulated less ROS, more proline and soluble sugars, higher peroxidase and catalase activities and increased expression of stress-responsive genes expression, leading to enhanced low temperature stress tolerance compared to control plants (Zhou et al., 2012). Hence, overexpression of *ZmMPK4* in transgenic tobacco results in increased tolerance to low temperature stress.

High temperature tolerance

Production of plants tolerant to high temperature stress is of immense significance in the light of global warming and climate change. Plant cells respond to high temperature stress by re-programming their genetic machinery for survival and reproduction (Grover et al., 2013). High temperature tolerance in transgenic plants has largely been achieved either by overexpressing heat shock protein genes or by altering levels of heat shock factors that regulate expression of heat shock and non-heat shock genes (Horváth et al., 2012).

When a new gene encoding an aldo-keto reductases (*ARKs*), which catalyses the conversion of glucose to sorbitol, from rice is expressed in tobacco plants, these transgenic plants show a better tolerance to high temperature stress (Grover et al., 2013). Turóczy et al. (2011) investigated two mechanisms how overexpression of ARKs in tobacco enhances high temperature tolerance. First, many AKRs are effective in the synthesis of osmolytes, thus having important role in the osmoregulation, which is an important process in plants for the acquisition of desiccation tolerance due to high temperature. Second, AKRs have also been shown to be effective in the detoxification of lipid peroxidation and/or glycolysis derived reactive carbonyls such as malondialdehyde 4-hydoxy-nonenal, methylglyoxal.

Over-expression of rice Zn-finger protein gene *ZFP177* in transgenic tobacco plants resulted in enhanced tolerance following high temperature stress (Grover et al., 2013). Huang et al. (2008) proposed that a possible role for ZFP177 in temperature stress tolerance is to down-regulate accumulations of some proteins that are not benefit to resistance to temperature stress by ubiquitinylating and targeting them for degradation Likewise, overexpression of pepper *WRKY40* in tobacco enhanced high temperature tolerance (Dang et al., 2013). Dang et al. (2013) investigated the mechanism how the overexpression of *WRKY40* enhanced tolerance to high temperature and they found that the overexpression of *WRKY40* increase transcript levels of hypersensitive response asccociated genes (*NtHSR20*, *NtHSR515*), ethylene biosynthesis related genes (*NtACS6*, *NtEFE26*), ROS detoxification associated genes (*NtSOD1, NtGST1* and *NtAPX*), as well as the heat-shock genes (*NtHSP18*, *NtsmallHSP* and *NtHSF2)*. These genes interact in different ways and improve tobacco tolerance to high temperature.

Heavy metals tolerances

Metal ions (Fe, Mn, Zn and Cu) are essential elements for plant growth and development (Han et al., 2013; Yadav, 2010). Heavy metal pollution is increasing rapidly due to excessive emission of industrial waste, wastewater irrigation, unreasonable pesticides and plastic sheeting (Wu et al., 2013). Though metal concentrations are increasing in soil and water, still there is a shortage of essential metals since they are found in unusable form for plants. Excess heavy metals in plant cells result lipid peroxidation (membrane deterioration), altered ion homeostasis and generation of ROS (Choe et al., 2013; Yadav, 2010) and adversely affected gene expression and slow photosynthesis (Wu et al., 2013). Heavy metal accumulation in plants is especially dangerous besides crop productivity reduction since it allows heavy metal contamination to food chains (Ovecka and Takac, 2014; Wu et al, 2013).

Iron

Iron is one of the essential micronutrients required by all plant and it is chelated by citric acid. However, in calcareous soil solution where the concentration of free Fe is far below normally required (10^{-15} M), the solubility of these ions is very poor (Han et al., 2013). Therefore, Fe-deficiency is a worldwide problem for crop production. Han et al. (2013) introduced *MxCS1* gene (isolated from *Malus xiaojinensis*) into tobacco, promoting the synthesis of citrate synthase and increasing citric acid levels. Overexpression of *MxCS1* improved Fe stress tolerance in transgenic tobacco and flowers presented morphological abnormalities. Higher concentrations of Fe, Mn, Cu and Zn in young leaves and flowers were found in transgenic plants as compared with WT plants. In addition to its role in long-distance metal transport, citric acid may be involved in the regulation of metal transfer within cells as well (Han et al., 2013).

Aluminum

Aluminum (Al) stress represses mitochondrial respiration and leads to increased ROS levels in plants (Panda et al., 2013). Mitochondrial alternative oxidase (AOX) uncouples respiration from mitochondrial ATP production. In order to study the function of tobacco AOX gene under Al stress, Panda et al. (2013) overexpressed *NtAOX1* under the control of the cauliflower mosaic virus (CaMV) 35 S promoter in sensitive *Nicotiana tabacum* L. cell lines. Endogenous *AOX1* gene expression and AOX protein levels were higher in transformed tobacco cell lines than in WT cell lines. They explained that a decreased respiratory inhibition and reduced ROS production with a better growth capability were the significant features that characterized AOX1 transformed

Table 2. Genes used to enhance tolerance to biotic stresses in tobacco.

Gene	Source	Role	Reference
WRKY40	Pepper	Enhanced pathogen tolerance	Dang et al. (2013)
BrERF11	Chinese cabbage	Resistance to disease by *Ralstonia solanacearum*	Lai et al., (2013)
ech42	*Trichoderma virens*	Resistance to fungal pathogens	Shah et al. (2011)
cry1Ac and *cry1Ab*	Artificially constructed	Resistance to *Spodoptera exigua* and *Helicoverpa armigera* larvae	Sohail et al. (2012)
Cadherin gene silencing	*Bacillus thuringiensis*	Alleviated *Manduca sexta* larvae resistance to cry toxin	Porta et al. (2011)

cell lines under Al stress. These results demonstrate that AOX plays a critical role in Al stress tolerance with an enhanced respiratory capacity, reducing mitochondrial oxidative stress burden and improving the growth capability in tobacco cells.

Cadmium

Cadmium (Cd) is considered an extremely significant pollutant due to its high toxicity to many organisms. Cadmium generates oxidative stress by interfering with antioxidant defense responses, which leads to the indirect production of oxygen-free radicals. These highly reactive species lead to multiple effects including membrane peroxidation, loss of ions, protein cleavage, enzyme inactivation and DNA strand damage (Khoudi et al., 2012). Plants have evolved several mechanisms to cope with Cd, the most important of which is vacuolar sequestration. Cadmium can be directly transported into vacuoles by cations/H+ exchangers, such as CAXs (calcium exchangers), which are energized by the pH gradient established by proton pumps. Khoudi et al. (2012) overexpressed a cDNA encoding wheat vacuolar H+-pyrophosphatase (TaVP1, a V-H-PPase) to invest-tigate whether this proton pump would enhance Cd tolerance. *TaVP1*-expressing plants were found to be more tolerant to Cd compared to WT plants when exposed to various concentrations of Cd. Despite the fact that *TaVP1*-transgenic plants had higher Cd content in their shoots and roots, *TaVP1*-transgenic plants performed much better than WT plants. Taken together, Khoudi et al. (2012) suggested that higher expression of a vacuolar proton pump contributed to Cd tolerance.

BIOTIC FACTORS

The biotic factors that greatly affect productivity of tobacco include viruses, fungi and bacteria, weeds and insects. As a result, the potential crop losses are estimated to be 30, 18 and 15% caused by weeds, pests and microorganisms (Sutherst et al., 2011). The classical method to protect the crop losses is extensive usage of chemical treatments but this is costly and bad for the environment. For instance, according to FAO (2005)

report, the world market for fungicides was estimated to be US$6 billion at the end-user level in 1999. However, overexpressing "resistant" genes has lead to promising results. Examples of such genes engineered into transgenic tobacco to enhance its defense under biotic stresses are shown in Table 2.

Microbial pathogens and viruses

Major losses of yield and deterioration in quality are the results of attack by an array of microbial pathogens especially fungi (FAO, 2005), and plants develop different mechanisms to counteract pathogens. Plants protection against infection by pathogens is achieved through the intervention of viral development or replication, called pathogen-derived resistance (PDR) (Koh et al., 2014). PDR of viruses can be divided into two categories: (1) RNA-mediated resistance that involves the transformation of a partial sequence of the virus genome into the plant, and (2) protein-mediated resistance that involves the transformation of the viral full-length protein-encoding gene into the plant. Besides PDR, Ding et al (2010) showed that RNA silencing serve as antiviral defense mechanisms.

Plant defense reactions are regulated by complicated signaling networks, which generally include ion fluxes across the plasma membrane, oxidative burst, changes in phytohormone and sugar levels, MAP kinase cascades and defense associated genes regulation by transcription factors (Lai et al., 2013; Bolouri and Van den Ende, 2013). For instance, overexpression of pepper *WRKY40* in tobacco strengthened *Ralstonia solanacearum* (a bacterial pathogen) tolerance (Dang et al., 2013). Similarly, overexpression of a Chinese cabbage *BrERF11* transcription factor enhanced disease resistance to *R. solanacearum* in tobacco (Lai et al., 2013). In addition, development of transgenic tobacco expressing an endochitinase (*ech42*) gene from the biocontrol fungus *Trichoderma virens* showed enhanced resistance to fungal pathogens (Shah et al., 2011).

Insects

Another major threat to crop production worldwide that

limit agricultural productivity are pests. The worldwide economic damage caused by insect pests to agricultural and horticultural crops and to orchards stands at a hundred billion dollars annually (FAO, 2005). Chemical insecticides are the most effective and still the predominant type of pest control strategies employed today. However, non-judicious and continuous use of these insecticides results in the development of resistance in insect populations against these insect-cides. Moreover, these have harmful effects on non-target/beneficial insects, potentially damaging natural environments. Insect resistant transgenic plants have emerged as potential alternatives to the synthetic insecticides. Genes encoding the insecticidal crystal protein were isolated from *Bacillus thuringiensis* (*Bt*) and successfully transformed into plants to produce insect-resistant plants (Porta et al., 2011; Sohail et al., 2012). *Bt* toxins are safe to non-target insects due to their specific activity towards target insects. Moreover, these toxins are not persistent in the environment unlike synthetic insecticides and can be used on large scales without any direct/indirect hazardous effects on humans.

However, transgenic plants with a narrow spectrum of insecticidal activity may not be protected against a wider array of insects. This problem was overcome by expressing more than one toxin simultaneously. Sohail et al. (2012) investigated a combination of toxins such as *cry1Ac* and *cry1Ab*, artificially constructed genes, to broaden the action spectrum in tobacco. They found 12% and 62% resistance to *Spodoptera exigua* and *Helicoverpa armigera* larvae. Both toxins showed synergistic effect in tobacco and broadened the spectrum of plant activity against insects.

Resistance to Cry toxins, produced by *B. thuringiensis,* has been linked with mutations in the *cadherin* gene (Porta et al., 2011; Sohail et al., 2012). One strategy effective to overcome larval resistance to Cry1A toxins is the production of Cry1AMod toxins that lack helix α-1since helix α-1is responsible for resistance (Porta et al., 2011). Cry1AMod are able to form oligomeric structures without binding to cadherin receptor and were shown to be toxic to cadherin-silenced *Manduca sexta* larvae and *Pectinophora gossypiella* strain with resistance linked to mutations in a cadherin gene (Porta et al., 2011). They found that Cry1AbMod protein produced in tobacco plants retains its functional toxic activity against susceptible and tolerant *M. sexta* larvae due to the silencing of *cadherin* receptor by RNAi.

CONCLUSION

The use of transgenes to improve the tolerance of tobacco to abiotic and biotic stresses remains an attractive option. Different genes contribute to specific traits for stress tolerance can be successfully engineered into tobacco and they improve tolerance dramatically. However, this review focuses on single gene transfers. In the future, multiple gene transfer might provide better enhancements to both abiotic and biotic stresses.

Conflict of Interests

The author(s) have not declared any conflict of interests

REFERENCES

Arif A, Zafar Y, Arif M, Blumwald E (2013). Improved growth, drought tolerance, and ultrastructural evidence of increased turgidity in tobacco plants overexpressing arabidopsis vacuolar pyrophosphatase (AVP1). Mol. Biotechnol. 54(2):379-392.

Bolouri MM, Van den Ende W (2013). Sweet immunity in the plant circadian regulatory network. J. Experimen. Botany 64(6):1439-1449.

Choe YH, Kim YS, Kim IS, Bae MJ, Lee EJ, Kim YH, Park HM, Yoon HS (2013). Homologous expression of γ-glutamylcysteine synthetase increases grain yield and tolerance of transgenic rice plants to environmental stresses. J. Plant Physiol. 170(6):610-618.

Dang FF, Wang YN, Yu L, Eulgem T, Lai Y, Liu ZQ, Wang X, Qiu AL, Zhang TX, Lin J, Chen YS, Guan DY, Cai HY, Mou SL, He SL (2013). CaWRKY40, a WRKY protein of pepper, plays an important role in the regulation of tolerance to heat stress and resistance to *Ralstonia solanacearum* infection. Plant Cell Environ. 36(4):757-774.

Ding SW (2010). RNA-based antiviral immunity. Nature reviews Immunol.10:632-644.

Doltchinkova V, Angelova P, Ivanova E, Djilianov D, Moyankova D, Konstantinova T, Atanassov A (2013). Surface electric charge of thylakoid membranes from genetically modified tobacco plants under freezing stress. J. Photochem. Photobiol. B: Biol. 119:22-30.

Eltelib HA, Fujikawa Y, Esaka M (2012). Overexpression of the acerola (*Malpighia glabra*) Monodehydroascorbate reductase gene in transgenic tobacco plants results in increased ascorbate levels and enhanced tolerance to salt stress. South Afr. J. Bot. 78:295-301.

FAO (2005). The status of research and application of crop biotechnologies in developing countries. Rome, Italy.

Gao Q, Guo QF, Xing SC, Zhao MR, Li F, Wang W (2007). The characteristics of expansins in wheat coleoptiles and their responses to water stress. J. Plant Physiol. Mol. Biol. 35:402-410.

Goel S, Madan B (2014). Genetic engineering of crop plants for abiotic stress tolerance IN: Ahmed P, Rasool S (Ed): Emerging Technologies and Management of Crop Stress Tolerance, Volume 1. Elsevier, Amsterdam.

Grover A, Mittal D, Negi M, Lavania D (2013). Review: Generating high temperature tolerant transgenic plants: achievements and challenges. Plant Sci. 205-206:38-47.

Han D, Wang L, Wang Y, Yang G, Gao C, Yu Z, Li T, Zhang X, Ma L, Xu X, Han Z (2013). Overexpression of *Malus xiaojinensis CS1* gene in tobacco affects plant development and increases iron stress tolerance. Scientia Horticulturae 150:65-72.

Horváth I, Glatz A, Nakamoto H, Mishkind ML, Munnik T, Saidi Y, Goloubinoff P, Harwood JL, Vigh L (2012). Heat shock response in photosynthetic organisms: Membrane and lipid connections. Prog. Lipid Res. 51:208-220.

Huang J, Wang MM, Jiang Y, Bao YM, Huang X, Sun H, Xu DQ, Lan HX, Zhang HX (2008). Expression analysis of rice A20/AN1-type zinc finger genes and characterization of *ZFP177* that contributes to temperature stress tolerance. Gene 420:135-144.

Hui-Ming G, Zhao-Chun L, Han Z, Yue-Zhi X, Hong-Mei C (2011). Cloning of cotton CBF gene for cold tolerance and its expression in transgenic tobacco. Acta Agron. Sin.37(2):286-293.

Keunen E, Peshev D, Vangronsveld J, Van den Ende W, Cuypers A (2013). Plant sugars are crucial players in the oxidative challenge during abiotic stress. Extending the traditional concept. Plant Cell Environ. 36(7):1242-1255.

Khoudi H, Maatar Y, Gouiaa S, Masmoudi K (2012). Transgenic tobacco plants expressing ectopically wheat H+-pyrophosphatase (H+-PPase) gene *TaVP1* show enhanced accumulation and tolerance

to cadmium. J. Plant Physiol. 169: 98- 103.

Koh KW, Lu HC, Chan MT (2014).Virus Resistance in Orchids. Plant Sci. http://dx.doi.org/10.1016/j.plantsci.2014.04.015.

Kumar D, Singh P, Yusuf MA, Upadhyaya CP, Roy SD, Hohn T, Sarin NB (2013). The *Xerophyta viscosa* Aldose Reductase (*ALDRXV4*) Confers Enhanced Drought and Salinity Tolerance to Transgenic Tobacco Plants by Scavenging Methylglyoxal and Reducing the Membrane Damage. Mol. Biotechnol. 54(2):292-303.

Lai Y, Dang, F, Lin J, Yu L, Shi Y, Xiao Y, Huang M, Lin J, Chen C, Qi A, Liu Z, Guan D, Mou S, Qiu A, He S (2013). Overexpression of a Chinese cabbage *BrERF11* transcription factor enhances disease resistance to *Ralstonia solanacearum* in tobacco. Plant Physiol. Biochem. 62:70-78.

Li F, Han Y, Feng Y, Xing Z, Zhao M, Chen Y, Wang W (2013). Expression of wheat expansin driven by the *RD29* promoter in tobacco confers water-stress tolerance without impacting growth and development. J. Biotechnol. 163:281-291.

Li F, Xing S, Guo Q, Zhao M, Zhang J, Gao Q, Wang G, Wang W (2011). Drought tolerance through over-expression of the expansin gene *TaEXPB23* in transgenic tobacco. J. Plant Physiol. 168:960-966.

Liu Q, Xu K, Zhao L, Pan, Y, Jiang B, Zhang H, Liu G (2011).Overexpression of a novel chrysanthemum NAC transcription factor gene enhances salt tolerance in tobacco. Biotechnol. Lett. 33:2073-2082.

Panda SK, Sahoo L, Katsuhara M, Matsumoto H (2013). Overexpression of alternative oxidase gene confers aluminum tolerance by altering the respiratory capacity and the response to oxidative stress in tobacco cells. Mol. Biotechnol. 54(2):551-563.

Ovecka M, Takac T (2014). Managing heavy metal toxicity stress in plants: Biological and biotechnological tools. Biotechnol. Advances 32:73-86.

Peshev D, Vergauwen R, Moglia A, Hideg E, Van den Ende W (2013). Towards understanding vacuolar antioxidant mechanisms: a role for fructans?. J. Experimen. Botany 64(4): 1025-1038.

Pilon-Smits EAH, Ebskamp MJM, Paul MJ, Jeuken MJW, Weisbeek PJ, Smeekens SCM (1995). Improved performance of transgenic fructan-accumulating tobacco under drought stress. Plant Physiol. 107:125-130.

Porta H, Jiménez G, Cordoba E, León P, Soberón M, Brao A (2011). Tobacco plants expressing the Cry1AbMod toxin suppress tolerance to Cry1Ab toxin of *Manduca sexta* cadherin-silenced larvae. Insect Biochem. Molecular Biol. 41:513- 519.

Shah MR, Mukherjee PK Eapen S(2011). Expression of a fungal endochitinase gene in transgenic tomato and tobacco results in enhanced tolerance to fungal pathogens. Physiol. Mol. Biol. Plants 16(1):39-51.

Singh AK, Kumar R, Pareek A, Sopory SK, Singla-Pareek SL (2012). Overexpression of Rice CBS Domain Containing Protein Improves Salinity, Oxidative, and Heavy Metal Tolerance in Transgenic Tobacco. Mol. Biotechnol. 52:205-216.

Sohail MN, Karimi SM, Asad S, Mansoor S, Zafar Y, Mukhtar Z (2012). Development of broad-spectrum insect-resistant tobacco by expression of synthetic *cry1Ac* and *cry2Ab* genes. Biotechnol. Lett. 34:1553-1560.

Sutherst RW, Constable F, Finlay KJ, Harrington R, Luck J, Zalucki MP (2011). Adapting to crop pest and pathogen risks under a changing climate. WIREs Clim. Change 2:220-237.

Turóczy Z, Kis P, Török K, Cserháti M, Lendvai A, Dudits D, Horváth GV (2011). Overproduction of a rice aldo-keto reductase increases oxidative and heat stress tolerance by malondialdehyde and methylglyoxal detoxification. Plant Mol. Biol. 75:399-412.

Van den Ende W, El-Esawe S (2014). Sucrose signaling pathways leading to fructan and anthocyanin accumulation: A dual function in abiotic and biotic stress responses? Environ. Exp. Bot. (ahead of print), art.nr. http://dx.doi.org/10.1016/j.envexpbot.2013.09.017.

Wang B, Zhang Q, Liu J, Li G (2011). Overexpression of *PtADC* confers enhanced dehydration and drought tolerance in transgenic tobacco and tomato: Effect on ROS elimination. Biochem. Biophys. Research Commun. 413:10-16.

Wu H, Shena Y, Hua Y, Tana S, Lin Z (2011). A phytocyanin-related early nodulin-like gene, *BcBCP1*, cloned from *Boea crassifolia* enhances osmotic tolerance in transgenic tobacco. J. Plant Physiol. 168:935-943.

Wu J W, Shi Y, Zhu Y X, Wang Y C Gong H J (2013). Mechanisms of enhanced heavy metal tolerance in plants by silicon: A review. Pedosphere 23(6):815-825.

Wu L, Zhou M, Shen C, Liang J, Lin J (2012). Transgenic tobacco plants over expressing cold regulated protein CbCOR15b from *Capsella bursa-pastoris* exhibit enhanced cold tolerance. J. Plant Physiol. 169:1408- 1416.

Yadav SK (2010). Heavy metals toxicity in plants: An overview on the role of glutathione and phytochelatins in heavy metal stress tolerance of plants. South Afr. J. Botany 76:167-179.

Yang S, Tang X, Ma N, Wang L Meng Q (2011). Heterology expression of the sweet pepper *CBF3* gene confers elevated tolerance to chilling stress in transgenic tobacco. J. Plant Physiol.168:1804-1812.

Yue Y, Zhang M, Zhang J, Duan L, Li Z (2011). *Arabidopsis LOS5/ABA3* overexpression in transgenic tobacco (*Nicotiana tabacum* cv. *Xanthinc*) results in enhanced drought tolerance. Plant Sci. 18:405-411.

Zhai Y, Wang Y, Li Y, Lei T, Yan F, Su L, Li X, Zhao Y, Sun X, Li Y, Wang Q (2013). Isolation and molecular characterization of *GmERF7*, a soybean ethylene-response factor that increases salt stress tolerance in tobacco. Gene 513:174-183.

Zhou S, Zhang Z, Tang Q, Lan H, Li Y, Luo P (2011). Enhanced V-ATPase activity contributes to the improved salt tolerance of transgenic tobacco plants overexpressing vacuolar Na$^+$/H$^+$ antiporter *AtNHX1*. Biotechnol. Lett. 33:375-380.

Zhou Y, Zhang D, Pan J, Kong X, Liu Y, Sun L, Wang L, Li D (2012). Overexpression of a multiple stress-responsive gene, *ZmMPK4*, enhances tolerance to low temperature in transgenic tobacco. Plant Physiol. Biochem. 58:174-181.

Full Length Research Paper

BengaSaVex: A new computational genetic sequence extraction tool for DNA repeats

OLUWAGBEMI, Oluseun Olugbenga[1,2]*, IMOLORHE, Samuel[2] and AGOZIE, Victor Okechukwu[2]

[1]Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, United States of America.
[2]Bioinformatics Unit, Department of Computer and Information Sciences, School of Natural and Applied Science, College of Science and Technology, P.M.B 1023 Covenant University, Ogun State, Nigeria, West Africa.

The scourge of infectious diseases is one of the problems contending with humanity. All infectious diseases are caused by pathogens. A major problem in biological research is the creation of enormous and redundant amounts of genomic data. From this large volume of generated data, biologists select a subset of each sequence known as DNA nucleotide subsequences "words", for extended scientific analysis. Computational biology aids this pruning process by providing computerized tools to generate vital information with biological significance from these data. This research aimed to develop new tools for extracting DNA repeats from the gene sequences and also to perform a comparative analysis with existing tools having similar or closely-related functions. We were able to develop *BengaSaVex* (GBenga Samuel Victor genetic sequence extraction tool) and provide a sequential *in-silico* genetic-sequence-filtering functionality to identify repeated DNA nucleotide subsequences within the genes of some microorganisms, evaluated the potential benefits and applications of identifying such repeated sequences, and finally, performed an *in-silico* comparative analysis between *BengaSaVex* and tandem repeat finder.

Key words: *BengaSaVex*, DNA, repetitive sequence, *in-silico* analysis, computational genomics.

INTRODUCTION

Over the years, biologists and computational biologists have conducted experiments related to the sequences of some pathogens and other micro organisms. One of the major problems in biological research is the creation of enormous and redundant amounts of genomic data from DNA sequencing projects performed (Baxevanis, 2003;

Wang and Zhang, 2005; Myers et al., 2006; Lathe et al., 2008; Oluwagbemi and Omonhinmin, 2008; Oluwagbemi, 2012). Biologists select a subset of each sequence also known as DNA nucleotide subsequences "words", for extended scientific analysis. Computational biology complements this pruning process by providing repeat

finding programs to help analyze and provide useful information about interesting words, with the assumption that under or over-represented words have significant biological functions.

The biological significance of DNA repeats cannot be underestimated. DNA repeats play a significant role in the biological sciences (Jurka, 1998). Transposable elements are hidden in many repetitive DNA sequences. Experimental research and analysis on these repetitive sequences can help reveal transposable elements that are associated with genomic evolution.

The aim of this research was to develop a useful extraction tool (*BengaSaVex*), for *in-silico* analysis on the gene sequences of some microorganisms. Some pathogens are only being used as an example of how the program works. The objectives of this research were: (i) to develop *in-silico* simultaneous genetic sequence-filtering tools for *in-silico* analysis, by using object-oriented programming languages in C++, (ii) to identify repeated DNA nucleotide subsequences within the genes of some microorganisms, (iii) to evaluate the potential benefits of (ii) and (iv) to conduct a comparative analysis between *BengaSaVex - C++ version* and tandem repeat finder (Benson, 1999).

The biological rationale for undertaking this research stems from the fact that prominent feature of DNA can be identified by the frequency with which repeated substrings exist. For instance, this seems to be true for eukaryotes (Lander et al., 2001). Some repeats have been found to aid the provision of structural mechanism (Huang et al.,1998), while others have been identified to affect bacterial virulence, among microbes which have the tendency to cause human infections (van Belkum et al., 1998). This makes a study on repeats a promising and interesting one.

In this paper, we devised a genetic subsequence extraction tool using the C++ programming language for its implementations. We named this tool as *BengaSaVex*. The tool has the capability to extract repetitive DNA sequences from a collection of multiple gene sequences of microorganisms including infectious-disease causing organisms; then estimate the relationship that exists between the lengths of extracted repeated sequence and the computational time taken to extract these repeated sequences. Insight gained from the analysis of these duplicated sequences could help accelerate the pace of research in this domain by causing a motivation for the development of more efficient tools, especially, since there is a huge volume of sequence data available.

Several traditional repeat finding programs have been developed and applied to different gene sequences. They are as described in Table 1.

In summary, this paper details the algorithm underlying the development of *BengaSaVex,* describes the mechanism of data collection, explores the potential benefits of identifying DNA repeats in gene sequences of computa-tional biology related research, presents the results generated by the new tools and its comparative analysis with some of the existing tools with similar or closely related functions (Saha et al., 2008).

MATERIALS AND METHODS

Data collection

Data for this research work was sourced from the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/) and also from the Sanger Institute (ftp://ftp.sanger.ac.uk/pub/pathogens/spn/). The sequence data of some microorganisms were sourced from various gene banks. Table 2 shows the sources of data used in the analysis. Each genome sequence data for respective organisms was simultaneously inserted into the input file of BengaSaVex.

Implementation

C++ programming language was used for the implementation of *BengaSaVex*. The multiple sequence data for different pathogens were stored inside an input file for BengaSaVex, for onward *in-silico* analysis. The input file (many.in.txt) contains multiple gene sequences of infectious disease-causing organisms to be analyzed, while the output file (many.out.txt) contains the results generated by BengaSaVex after running the executable version of the software (BengaSaVex.exe). BengaSaVex was developed using algorithms to compare sub-strings of gene sequences that are identical within genome sequence of pathogens as shown in (List 1). The algorithm depicted below shows its operations on repeat sequences.

List 1: *BengaSaVex Algorithm*
```
    Begin
        Input S1 ,………….., Sm: the m set of pathogen gene
sequence
                While (!end of file) do
                    Get next set of gene sequence
                        for all i=1 to n do
                            function Search and Compare
subsets of gene sequence S11 with S12 within S1,…… until S1n
                                Identify        repeated
sequences from S1,……Sm
                                Output repeats R1,……Rm each for
Sequence S1,…..,Sm
                        end for
                        Output frequencies Rf1, Rf2, Rf3,…………,
Rfm for each repeat
        Compute corresponding time (T1,……,Tm) to search and
extract each repeat
        Return S1,……,Sm; frequencies Rf1, Rf2, Rf3,……Rfm for
each repeat; time (T1,……,Tm) to search and extract each repeat
    End
```

RESULTS

BengaSaVex has the capability to perform sequential *in-silico* analysis on hundreds to thousands of large genome sequences. However, for the purpose of this manuscript, we only analyzed close to 15 large genome sequences. We present the results of eight of them as produced by *BengaSaVex*, based on *in-silico* analysis performed on the gene sequences of some organisms as shown in Table 2. Some of the repeats were found to be intergenic. We also provide the results of a comparative analysis of *BengaSaVex* with the tandem repeat finding program (Table 3).

Table 1. Tabulated literature review of some traditional repeat finding programs.

Related work	Description and reference
RepeatMasker	**RepeatMasker**, a prominent software, was developed to identify, classify and mask repetitive gene sequences. RepeatMasker finds repetitive sequence by performing an alignment of the input sequence against a library of known repeats (Smit and Green, 2002; Tarailo-Graovac and Chen, 2009).
RepeatScout	**RepeatScout** was another program developed to identify repetitive sequence in large genomic sequence (Price et al., 2005).
SAGRI	**SAGRI** (Spectrum Assisted Genomic Repeat Identifier), was a tool developed as a novel approach to detecting repeats in genomic sequences. SAGRI performs a double scan on the genome sequence (Do et al., 2008). It's a tool that was developed to efficiently locate possible ancient repeats in genomic sequences produced encouraging results (Singh et al., 2007).
RECON	**RECON**, an automated software for identifying repetitive sequences of newly sequenced genomes, was also developed (Bao and Eddy, 2002).
WindowMasker	**WindowMasker** was developed to identify and mask highly repetitive subsequences in the DNA sequence of a genome (Morgulis et al., 2006).
RepeatFinder	Algorithms such as **RepeatFinder** (Volfovsky et al., 2001) are also useful in *in-silico* analyses.
RepeatGluer	**RepeatGluer** (Pevzner et al., 2004)
PILER	Recently, **PILER** (Edgar and Myers, 2005) have increasingly automated the identification of repeat families from genomic sequence
ReAs	**ReAs** algorithm was applied in recovering ancestral sequences from transposable elements (Li et al., 2005).
REPuter	**REPuter** (http://bibiserv.techfak.uni-bielefeld.de/reputer/) , another repeat finding program, was developed by Kurtz and colleagues (Kurtz et al., 2001).
Dst	**Dst** (http://alce.med.umn.edu/newdst.html; Virtual Genome Center, unpublished), is another repeat finding program.
REPRO	**REPRO**, another program, helps to identify repeats in gene sequences of proteins [http://mathbio.nimr.mrc.ac.uk/~rgeorge/repro; (George and Heringa, 2000)].
RepeatAround	**RepeatAround** software was a repeat finding tool created by (Goios et al., 2006) - http://portugene.com/repeataround.html).
OMWSA	The **OMWSA** is another online tool for repeat finding and visualization (Du, 2007).
REPFIND	**REPFIND** online repeat finding tool (Betley et al., 2002), (http://zlab.bu.edu/repfind/form.html) was created by Bentley and colleagues.
Tandem Repeat Finder	**Tandem Repeat Finder** is yet another repeat finding program (Benson, 1999).

BengaSaVex - C++ version was used for this analysis, because it provided extraction time (in milliseconds) for the frequency of each direct repeated sequence. Analysis was performed on whole genome sequences of *Pseudomonas fluorescens* (Von Graevenitz and Weinstein, 1971; Picot et al., 2001), Hippea maritime DSM 10411 (Miroshnichenko et al., 1999), *Bartonella tribocorum* CIP 105476 (Heller et al., 1998), *Sinorhizobium meliloti* BL225C(Audic et al., 2009), *Brucella pinnipedialis* B2/94 (Whatmore, 2009; Audic et al., 2011), and *Staphylococcus aureus* [EMRSA15](methicillin-resistant strain) (Meier et al., 2001; Gordon and Lowy, 2008; Löffler

Table 2. *In-silico* analysis result from *BengaSaVex.*

Organism's sequence	DNA	References/accession number	Frequency of repeat	*BengaSaVex* -C++ version search and extraction time (s)
Bartonella tribocorum 105476	CIP	(Chomel and Piemont,1998) NCBI Reference Sequence: NC_010161.1 GI:163867306	Words with the maximum frequency (2) in the text are: CCGCCGCTGCTTTATTGATTAAACCCAGACAATTTC TAACTCGCCATAAGGAGAGAGACGTTCAGTATAG AGATATGACCACAAGCGACAGCCTTGCCAAAAACC CTTGCGTTTTCATAAACGTGACCATTTTTAAACAC TGATTATCATCACAAATATCATTCGTAATTTTTTCAC GGCTATTAACCACGGCATTACCATAGATTTGTA CATCATGATCAATATTGGATTTACCCATAGACCTTTG CATTTTCATAAATTCTTGCAAAGGACCCCACCCG AGCCGAGCCAGAAACCTTTGCATCATCAAAAATAC TTGCAAAACTCTTAATCCAAGCATTACAAAAGACA TGGGCATTGCCATAAATACACGCCCCCCCATAAAT GCGTGATTTACCATAAACATGGGAATTCCCATAAA AGCATTATCACAAACCATGGCATTGCCATAAACTAG ACCACAGATCCTAGCATTGTTACTAATTTTTGCA AATGCCCCCGCCTTAACATCATCAAAATCTCTTAAA GCACGAATGCGGTGTACAGTAATCCCTTCAAAAG TTGAACGTTGAAACTGTTAGATTACCCACGAGGCG TGAGCCAGTTTTGCCAGGACGGTTTTACCGGATCT CTTGGAAACCCCTTGGAAACCCCTTGGAAACCCCT TGGAAACCCCTTGGAAACCCCTTGGAAACCCCTTG GAAACCCCTTGGAAACCCCTTGGAAACCCCTTGGA AACCCCTTGGAAACCCCTTGGAAACCCCTTGGAAA CCCCTTGGAAACCCCTTGGAAACCCCTTGGAAACC CCTTGGAAACCCCTTGGAAACCCCTTGGAAACCCC TTGGAAACCCCTTGGAAACCCCTTGGAAACCCCTT GGAAACCCCTTGGAAACCCCTTGGAAACCCCTTGG AAACCCCTTGGAAACCCCTTGGAAACCCCTTGGAA ACCCCTTGGAAACCCCTTGGAAACCCCTTGGAAAC CCCTTGGAAACCCCTTGGAAACCCCTTGGAAACCC CTTGGAAACCCCTTGGAAACCCCTTGGAAACCCCT TGGAAACCCCTTGGAAACCCCTTGGAAACCCCTTG GAAACCCCTTGGAAACCCCTTGGAAACCCCTTGGA TGTAGGAAACTGTAGGAAACTGTAGGAAACTGTAG GAAACTGTAGGAAACTGTAGGAAACTGTAGGAAAC TGCGACTGCGACTGCGACTGCGACTGCGACTGCG ACTGCGACTGCGACTGCGACTGCGACTGCGACTG CG ACTGCGACTGCGACTGCGACTGCGACTGCGACTG CGACTGCGACTGCGACTGCGACTGCGACTGCGAC TG	35.637
BORRELIA afzelii Pko complete genome sequence		GI: 384206106 NCBI Reference Sequence: NC_017227.1	Nil	0.496

Table 2. Contd.

		Words with the maximum frequency (2) in the text are:	
Sinorhizobium meliloti BL225C complete genome	NCBI Reference Sequence: NC_017322.1	CCGCTTGTCCCTTCTCCCCGCCTGCGGGGAGAAGGT GCCGGCAGGCGGATGAGGGGCGGGGTGACCGCTAA GATTCCTTTTTCTGCGCAAATCAGATTCACCATTCAGG CCGGTGAAGGAGGCCGGCCTTTATGGCGAGAC CTTTTTCGAATGATCTTCGGGAACGCGTTGTCGATGC GGTGACGGGCGAGGGCCTATCGTGCCGGGCAGC GGCCAAGCGCTTCGGCATCGGCATCAGCACCGCGAT CGATTGGGTGCGGCGGTTTCGCGAGACGGGCAGC GCCGCACCCGGCCAGATGGGTGGGCACAAGCCCCG CAAGCTTTCCGGTCCGCACCGGGCTTGGCTGCTTT GCCGCTGCCGCGAGCGCGACTTCACGCTGCACGGAC TTGTCGCCGAGTTGAGCGAGCGCGGCCTGAAGGT GGATTATCGCGCCGTCTGGACCTTCGTGCACGAAGAG GGGTTGAGTTATAAAAAAAGACGCTGGTCGCCA GCGAACGGGAGCGGCCCGACGTCGCCCGCCACCGG GCACGATGGCTGAAGCACTGCCCCGGAATTGATCC AGCCGCCGGCAGTGGGCGATTTGTGCAAAACCCTTC GGGCGGCCCATATTGCGGTGCCTTGTCGCGAAAA ATCCGGCTTGCAGGCGGACGGCCTGCGGCGCCGGAT TTTCCACGAAAGTCCCTCGCAATTTGGGCCGTCA	134.556
Brucella pinnipedialis B2/94 chromosome 1, complete sequence	NC_015857.1 NCBI Reference Sequence: NC_015857.1	Words with the maximum frequency (2) in the text are: AATGCAGCGCACTGGCGCGATCTGCCTGCGACCTTC GGCAAATGGACAGCGGTTCATGCCCGCTTTCGGC:: :GCTGGTCGCACGCCGGTGTATGGGAAAGGCTTTTCC ATGCCCTGGCTGATACGCCGGACTTTGAATATGT:: CCTCATTGATAGCACCATATCGAAAGTCCACGCAGAT GCGGCGGGCGCAAAAGGGGGGCTGAAGCTGCCT: ::GCATCGGTCGCTCGCGCGGTGGATTGACGACCAAG CTGCATGCTGTTGTCGATGCTATCGGCCTACCGCT:: ::GCGAATAAAGCCAACACCCGGCCATTATGGTGACTG TCCGCAAGCTTCAAGCCTTCTATCCGGCTTAGAG: ::TGGATGGCTGCCAATGCAGCGCACTGGCGCGATCT GCCTGCGACCTTCGGCAAATGGACAGCGGTTCATG: ::CCCGCTTTCGGCGCTGGTCGCACGCCGGTGTATGG GAAAGGCTTTTCCATGCCCTGGCTGATACGCCGGA:: ::CTTTGAATATGTCCTCATTGATAGCACCATATCGAAA GTCCACGCAGATGCGGCGGGCGCAAAAGGGGGG ::CTGAAGCTGCCTGCATCGGTCGCTCGCGCGGTGGA TTGACGACCAAGCTGCATGCTGTTGTCGATGCTAT:: ::CGGCCTACCGCTGCGAATAAAGCCAACACCCGGCCA TTATGGTGACTGTCCGCAAGCTTCAAGCCTTCTA:: ::TCCGGCTTAGAGGGTGTGGGGCATGTCATTGCTGAT GCGGCCTATGATGCCGATCACTTAAGGGCCTTCA:: ::TTGCCAGCAATCTCAAGGCAACGGCTCAGATCAAGG CCAATCCAACACGTTCCAGTGTCCCAACAATCGA: ::CTGGAGGCTGTACAAGGAACGCCATCAGATTGAATG CTTTTTTTAACAAGTTGAAACGCTATCGTCGTATT::	26.428

Table 2. Contd.

Pseudomonas fluorescens SBW25 complete genome	GI:229587578 NCBI Reference Sequence: NC_012660.1	Nil	400.238
Staphylococcus (methicillin-resistant)	EMRSA-15 genome ftp://ftp.sanger.ac.uk/pub/pathogens/sa/	Words with the maximum frequency (3) in the text are: ::tttaacttaagttattagagcctcttatgcagttgctcagtcaactgtataccttt tgac::	124.688
Staphylococcus aureus strains-Epidemic EMRSA-16lineage	MRSA252.dna ftp://ftp.sanger.ac.uk/pub/pathogens/sa/	: Nil	1.351
Staphylococcus aureus- Highly transmissible MRSA sequence type(ST) 239 by MLST	EMBL/GenBank databases with accession number FN433596. ftp://ftp.sanger.ac.uk/pub/pathogens/sa/	Nil	0.619

Table 3. *In-silico* comparative analysis between ***BengaSaVex*** and some repeat finding programs (with respect to time only).

Sequence	*BengaSaVex* (s)	Tandem repeat finder (s)
Bartonella tribocourm CIP 105476	35.637	60.15
BORRELIA afzelii Pko NCBIReference Sequence: NC_017227.1	0.496	0.544
Sinorhizobium meliloti BL225C complete genome	134.556	65.12
Brucella pinnipedialis B2/94 complete genome	26.428	41.96
Staphylococcus (methicillin- resistant)	124.688	271.36
Staphylococcus aureus strains- Epidemic EMRSA-16lineage	1.351	21.62
Staphylococcus aureus MSSA476- methicillin-sensitive strain	95.855	216
Staphylococcus aureus- Highly transmissible MRSA sequencetype(ST) 239 by MLST	0.619	4.95

et al., 2010), *Staphylococcus aureus* [Epidemic EMRSA-16 lineage], *Staphylococcus aureus* [MSSA476-methicillin-sensitive strain], *Staphylococcus aureus* [highly transmissible MRSA sequence type(ST) 239 by MLST(TW20) and the *Heamophilus Influenza*. Their respective accession numbers were provided in the following section. These results (Tables 2 and 3) show that BengaSaVex can be used as a complementary tool with other existing repeat finding programs. REFIND did not work on long sequences, and so was not included in Table 3.

BengaSaVex GUI shows the functionalities of the tool for input-ting data, analyzing, outputting results of extracted repeats, frequency of extracted repeats, and time taken to extract the repeats (Figure 1).

DISCUSSION

Results produced show that *BengaSaVex* can be used as a complementary tool for repeat finding related researches. Research on repeated sequences can help
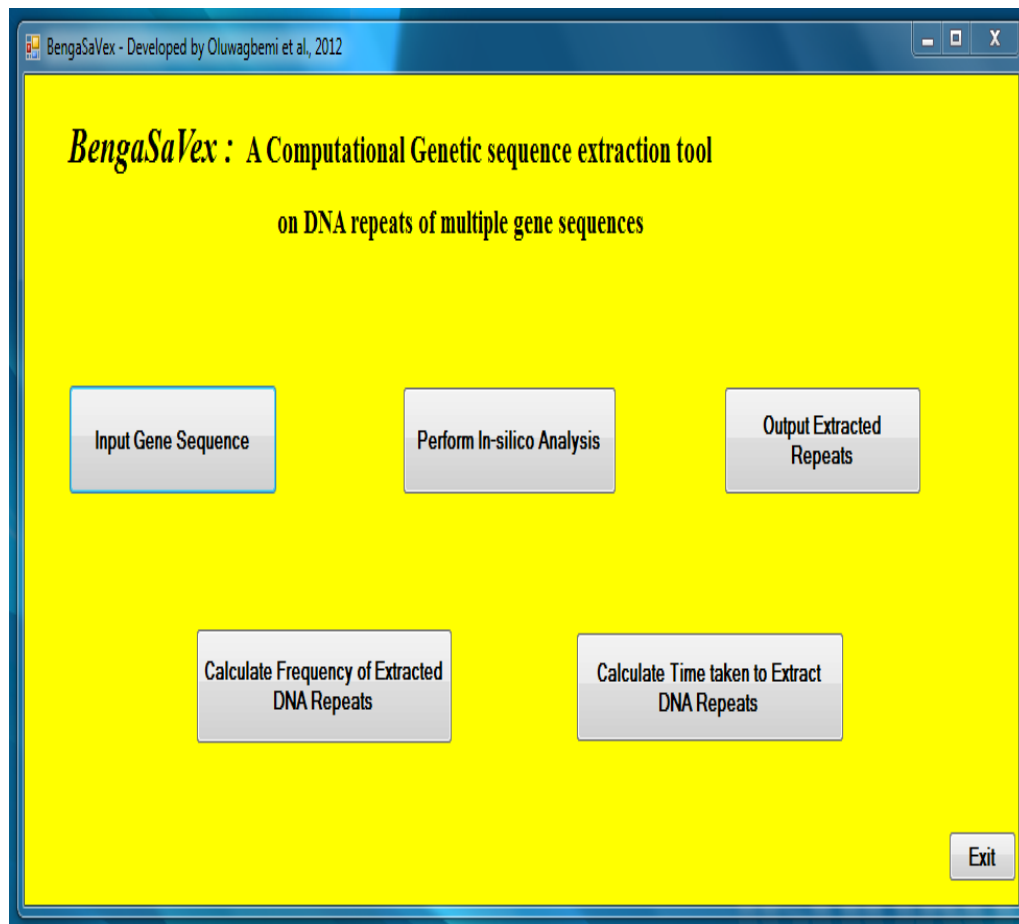
Figure 1. Graphical User Interface design of *BengaSaVex.*

provide interesting discoveries in the study of polymorphic patterns. Understanding the relationship between redundant gene filtering algorithms, programs and the corresponding genetic sequence they process, can help provide insight to developing programs with increased efficiency in carrying out this pruning process. This in turn, will help hasten or speed up the pace of research on DNA repeats, duplicated regions, sequence alignments and redundant genetic sequences of organisms and useful medicinal plants.

BengaSaVex has an added advantage to extract repeat sequences from multiple gene sequences of organisms, of which pathogens' are just one of the sample data. *BengaSaVex* also provides the corresponding frequencies of extracted sequences, and the time taken. *BengaSaVex* finds repeats in gene sequence of organisms.

Multifaceted applications of repeat analysis

Computational analysis finds expression in the processing of DNA repeats. Scientific research has found that DNA repeats help enhance flexibility in genetic and phenotypic features of pathogens and microorganisms (van Belkum et al., 1998). Variability in DNA repeats could help provide information about functional and evolutionary information on genetic diversity of such organisms (van Belkum, 1999a). Van Belkum as well as Delihas (van Belkum et al., 1999; Delihas, 2011), discovered and revealed the vital role sequence repeats play with the regulation of microbial gene expression. The significance of sequence repeats in epidemiologic typing cannot be underestimated (van Belkum, 1999b). Sequence repeats were also detected in *Escherichia coli'* sequence (Gur-Arie et al., 2000). Other scientists identified the potentials of DNA repeats in detecting certain virulent genes in pathogenic bacteria such as *H. influenza* (Hood et al., 1996; Power et al., 2009). Jansen and colleagues conducted an in-depth research on prokaryotes by detecting genes that are related to DNA repeats (Jansen et al., 2002; Treangen et al., 2009). Other scientists, such as Godde and Bickerton conducted similar experiments (Godde and Bickerton, 2006). Other related works that have been done in this regard are those

those of Cui as well as Bolotin (Cui et al., 2008; Bolotin et al., 2005). The application of DNA repeats have been emphasized in various infectious disease research over the years. Several functions of repeated sequences in MYCOPLASMA genomes have been highlighted in some studies (Ruland et al., 1990; Himmelreich et al., 1996; Himmelreich et al., 1997; Altshuler et al., 2000; Chambaud et al., 2001; Jaffe et al., 2004; Minion et al., 2004; Mrázek, 2006; Kassai-Jáger et al., 2008; Ma et al., 2008; Ma et al., 2012). DNA sequence repeats have also been found in enteric pathogens that are responsible for bacillary dysentery in humans (Jin et al., 2002; Wei et al., 2003; Yang et al., 2003; Phalipon and Sansonetti, 2007; Saurabh et al., 2011; Sun et al., 2011). Other studies have also revealed the significance of conducting comparative analyses and repeats in the genomes of various organisms (Powell et al., 1996; Chen et al., 2003; Ju et al., 2005; Rahim, 2008; Shikano et al., 2010; Labbe et al., 2011; Saker et al., 2011; Tyagi et al., 2011). Another study characterized repeats within sequences of exclusively prokaryotic genomes (Coenye and Vandamme, 2005).

A study has also shown the significance of repeated sequence in proteins and their relevance in network evolution (Hancock and Simon, 2005). Repeated sequences have the tendency of modifying other gene data to which they are associated, thus having the tendency of playing a role in the generation of genetic variation that underlies adaptive evolution (Kashi et al., 1997; Kashi and King, 2006). As stated above- genetic disorders do not cause disease; disease is defined as caused by an infectious agent (Clancy and Shaw, 2008). Research related to duplicated regions within gene sequences of microorganisms is of paramount interest in the field of computational biology and bioinformatics (Petes and Hill, 1988; Andersson and Hughes, 2009). Gene duplication has been found to be responsible for evolutionary mechanisms (Zhang, 2003). Duplicated regions in some organisms' chromosomes have also been found to play host to essential genes (Hillyard and Redd, 2007). Duplicated regions within the sequences of microorganisms like bacteria, play a significant role in their adaptation (Anderson and Roth, 1977). Scientists have also highlighted the relevance of duplicated regions within the sequence of certain pathogens (Larsson et al., 2005).

Conclusion

We developed *BengaSaVex* (a computational biology/-bioinformatics tool) for identifying and extracting repeats in gene sequences. This tool will complement other existing repeat finding tools to provide support for biological research. Future work on *BengaSaVex* is to improve the efficiency and also develop an online version.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

Altshuler D, Daly M, Kruglyak L (2000). Guilt by association. Nat. Genet. 26:135-137.
Anderson RP, Roth JR (1977). Tandem Genetic Duplications in Phage and Bacteria. Annu. Rev. Microbiol. 31: 473-505.
Andersson DI, Hughes D (2009). Gene Amplification and Adaptive Evolution in Bacteria. Annu. Rev. Genet. 43:167-195.
Audic S, Lescot M, Claverie J, Cloeckaert A, Zygmunt MS (2011). The genome sequence of Brucella pinnipedialis B2/94 sheds light on the evolutionary history of the genus Brucella. BMC Evol. Biol. 11:200.
Audic S, Lescot M, Claverie J, Scholz HC (2009). Brucella microti: the genome sequence of an emerging pathogen. BMC Genomics 10:352.
Bao Z, Eddy SR (2002). Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 12:1269-1276.
Baxevanis AD (2003). The Molecular Biology Database Collection: 2003 update. Nucleic Acids Res. 31(1):1-12.
Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27(2):573-80.
Betley JN, MC Frith, JH Graber, S Choo, JO Deshler (2002). A ubiquitous and conserved signal for RNA localization in chordates. Curr. Biol. 12:1756-1761.
Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151(8):2551-2561.
Chambaud I, Heilig R, Ferris S, Barbe V, Samson D, Galisson F, Moszer I, Dybvig K, Wróblewski H, Viari A, Rocha EPC, Blanchard A (2001). The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis, Nucleic Acids Res. 29(10):2145-2153.
Chen CY, Wu KM, Chang YC, Chang CH, Tsai HC, Tsai-Liao L, Liu Y, Chen H, Shen AB, Li J, Su T, Shao C, Lee C, Hor L, Tsai S (2003). Comparative genome analysis of vibrio vulnificus, a marine Pathogen. Genome Res. 13: 2577-2587
Clancy S, Shaw K (2008). DNA deletion and duplication and the associated genetic disorders. Nat. Educ. 1(1):23.

Coenye T, Vandamme P (2005). Characterization of mononucleotide repeats in sequenced prokaryotic genomes. DNA Res. 12(4):221-33.

Cui Y, Li Y, Yan Y, Yang R (2008). Clustered regularly interspaced short palindromic repeats: structure, function and application-a review. Wei Sheng Wu Xue Bao 48(11):1549-1555.

Delihas N (2011). Impact of Small Repeat Sequences on Bacterial Genome Evolution. Genome Biol. Evol. 3:959-973.

Do HH, Kwok PC, Franco PP, Wing KS, Louxin Z (2008). Spectrum-Based De Novo Repeat Detection in Genomic Sequences. J. Comput. Biol. 15(5):469-488.

Edgar RC, Myers EW (2005). PILER: identification and classification of genomic repeats. Bioinformatics 21(1):i152-i158.

George RA, Heringa J (2000). The REPRO server: finding protein internal sequence repeats through the web.Trends Biochem. Sci. 25:515-517.

Godde JS, Bickerton A (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J. Mol. Evol. 62(6):718-729.

Gordon RJ, Lowy FD (2008). Pathogenesis of Methicillin-Resistant Staphylococcus aureus Infection. Clin. Infect. Dis. 46(5):S350-S359.

Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM, Kashi Y (2000). Simple sequence repeats in Escherichia coli: abundance, distribution, composition, and polymorphism. Genome Res. 10(1):62-71.

Hancock JM, Simon M (2005). Simple sequence repeats in proteins and their significance for network evolution. Gene 345(1):113-118.

Heller R, Riegel P, Hansmann Y, Delacour G, Bermond D, Dehio C, Lamarque F, Monteil H, Chomel B, Piemont Y (1998). Bartonella tribocorum sp. nov., a new Bartonella species isolated from the blood of wild rats. Int. J. Syst. Evol. Microbiol. 48(4):1333-1339.

Hillyard DR, Redd MJ (2007). Identification of essential genes in bacteria. Methods Enzymol. 421:21-34.

Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li B, Herrmann R (1996). Complete Sequence Analysis of the Genome of the Bacterium Mycoplasma Pneumoniae. Nucleic Acids Res. 24(22):4420-4449.

Himmelreich R, Plagens H, Hilbert H, Reiner B, Herrmann R (1997). Comparative analysis of the genomes of the bacteria Mycoplasma pneumoniae and Mycoplasma genitalium. Nucleic Acids Res. 25(4):701-712.

Hood DW, Deadman ME, Jennings MP, Bisercic M, Fleischmann RD, Venter JC, Moxon ER (1996). DNA repeats identify novel virulence genes in Haemophilus influenzae. Proc. Natl. Acad. Sci. USA. 93(20):11121-11125.

Huang C, Lin Y, Yang Y, Huang S, Chen C (1998). The telomeres of Streptomyces chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. Mol. Microbiol. 28:905-916.

Jaffe JD, Stange-Thomann N, Smith C, DeCaprio D, Fisher S, Butler J, Calvo S, Elkins T, FitzGerald MG, Hafez N, Kodira CD, Major J, Wang S, Wilkinson J, Nicol R, Nusbaum C, Birren B, Berg HC, Church GM (2004). The Complete Genome and Proteome of Mycoplasma mobile. Genome Res. 14:1447-1461.

Jansen R, Embden JD, Gaastra W, Schouls LM (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol.

Jin Q, Yuan Z, Xu J, Wang Y, Shen Y, Lu W, Wang J, Liu H, Yang J, Yang F, Zhang X, Zhang J, Yang G, Wu H, Qu D, Dong J, Sun L, Xue Y, Zhao A, Gao Y, Zhu J, Kan B, Ding K, Chen S, Cheng H, Yao Z, He B, Chen R, Ma D, Qiang B, Wen Y, Hou Y, Yu J (2002). Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157. Nucleic Acids Res. 30(20):4432-4441.

Ju Z, Melissa C. Wells MC, Martinez A, Hazlewood L, Walter RB (2005). An in silico mining for simple sequence repeats from expressed sequence tags of zebrafish, medaka, Fundulus and Xiphophorus. In Silico Biol. 5: 0041.

Jurka J (1998). Repeats in genomic DNA: mining and meaning. Curr. Opin. Struct. Biol. 8(3):333-337.

Kashi Y, King D, Soller M (1997). Simple sequence repeats as a source of quantitative genetic variation. Trends Genet. 13(2):74-78.

Kashi Y, King DG (2006). Simple sequence repeats as advantageous

mutators in evolution. Trends Genet. 22(5):253-9.

Kassai-Jáger E, Ortutay C, Tóth G, Vellai T, Gáspári Z (2008). Distribution and evolution of short tandem repeats in closely related bacterial genomes. Gene 410(1):18-25.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29(22):4633-4642.

Labbé J, Murat C, Morin E, Le Tacon F, Martin F (2011). Survey and analysis of simple sequence repeats in the Laccaria bicolor genome, with development of microsatellite markers. Curr. Genet. 57(2):75-88.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, et al. (2001), Initial Sequencing and analysis of the human genome. Nature 409:860-921.

Larsson P, Oyston PCF, Chain P, Chu MC, Duffield M, Fuxelius H, E Garcia, G Hälltorp, D Johansson, KE Isherwood, PD Karp, E Larsson, Y Liu, S Michell, J Prior, R Prior, S Malfatti, A Sjöstedt, K Svensson, N Thompson, L Vergez, JK Wagg, BW Wren, LE Lindler, SGE Andersson, M Forsman, Titball RW (2005). The complete genome sequence of Francisella tularensis, the causative agent of tularemia. Nat. Genet. 37:153-159.

Lathe W, Williams J, Mangan M, Karolchik D (2008). Genomic Data Resources: Challenges and Promises. Nat. Educ. 1:3.

Li R, Ye J, Li S, Wang J, Han Y, et al. (2005). ReAS: Recovery of Ancestral Sequences for Transposable Elements from the Unassembled Reads of a Whole Genome Shotgun. PLoS Comput. Biol. 1(4):e43.

Löffler B, Hussain M, Grundmeier M, Brück M, Holzinger D, et al., (2010). Staphylococcus aureus Panton-Valentine Leukocidin Is a Very Potent Cytotoxic Factor for Human Neutrophils. PLoS Pathog. 6(1):e1000715.

Ma L, Jensen JS, Mancuso M, Hamasuna R, Jia Q, McGowin CL, Martin DH (2012). Variability of trinucleotide tandem repeats in the MgPa operon and its repetitive chromosomal elements in Mycoplasma genitalium. J. Med. Microbiol. 61(2):191-197.

Ma L, Taylor S, Jensen JS, Myers L, Lillis R, Martin DH (2008). Short tandem repeat sequences in the Mycoplasma genitalium genome and their use in a multilocus genotyping system. BMC Microbiol. 8:130.

Meier SP, Entenza JM, Vaudaux P, Francioli P, Glauser MP, Moreillon P (2001). Study of Staphylococcus aureus Pathogenic Genes by Transfer and Expression in the Less Virulent Organism Streptococcus gordonii. Infect. Immun. 69(2):657-664. Microbiol. 43(6):1565-75.

Minion FC, Lefkowitz EJ, Madsen ML, Cleary BJ, Swartzell SM, Mahairas GG (2004). The genome sequence of mycoplasma hyopneumoniae strain 232, the agent of swine mycoplasmosis. J. Bacteriol. 186(21):7123-7133.

Miroshnichenko ML, Rainey FA, Rhode M, Bonch-Osmolovskaya EA (1999). Hippea maritima gen. nov., sp. nov., a new genus of thermophilic, sulfur-reducing bacterium from submarine hot vents. Int. J. Syst. Evol. Microbiol. 49(3):1033-1038.

Morgulis A, Gertz EM, Schäffer AA, Agarwala R (2006). WindowMasker: window-based masker for sequenced genomes. Bioinformatics 22 (2):134-141.

Mrázek J (2006). Analysis of distribution indicates diverse functions of simple sequence repeats in Mycoplasma genomes. Mol. Biol. Evol. 23(7):1370-85.

Myers CL, Barrett DR, Hibbs MA, Huttenhower C, Troyanskaya OG (2006). Finding function: evaluation methods for functional genomic data. BMC Genomics 7:187.

Oluwagbemi OO, Omonhinmin CO (2008). Evaluating the Relationship Between Running Times and DNA Sequence Sizes using a Generic-Based Filtering Program. Pac. J. Sci. Technol. 9(2):656-666.

Oluwagbemi, OO (2012). Development of a prototype hybrid-grid-based computing framework for accessing bioinformatics databases and resources. Sci. Res. Essays 7(7):730-739.

Petes TD, Hill CW (1988). Recombination between repeated genes in microorganisms. Annu. Rev. Genet. 22:147-168.

Pevzner PA, Tang H, Tesler G (2004). De novo repeat classification and

fragment assembly. Genome Res. 14:1786-1796.

Phalipon A, Sansonetti PJ (2007). Shigella's ways of manipulating the host intestinal innate and adaptive immune system: a tool box for survival? Immunol. Cell Biol. 1-11.

Picot L, Abdelmoula SM, Merieau A, Leroux P, Cazin L, Orange N, Feuilloley MG (2001). Pseudomonas fluorescens as a potential pathogen: adherence to nerve cells. Microbes Infect. 3(12):985-995.

Powell W, Machray GC, Provan J (1996). Polymorphism revealed by simple sequence repeats. Trends Plants 1(7):215-222.

Power PM, Sweetman WA, Gallacher NJ, Woodhall MR, Kumar GA, Moxon ER, Hood DW (2009). Simple sequence repeats in Haemophilus influenza. Infect. Genet. Evol. 9(2):216-228.

Price AL, Jones NC, Pevzner PA (2005). De novo identification of repeat families in large genomes. Bioinformatics 21(Suppl. 1):351-358.

Rahim F(2008). *In silico* comparison of simple sequence repeats in high nucleotides-rich genomes of microorganism. Pak. J. Biol. Sci. 11(20):2372-2781.

Ruland K, Wenzel R, Herrmann R (1990). Analysis of three different repeated DNA elements present in the P1 operon of Mycoplasma pneumoniae: size, number and distribution on the genome. Nucleic Acids Res. 18(21):6311-6317.

Saker MM, Mohamed AA, Aly AA (2011). Comparative analysis of transformed potato microtubers and its non-transformed counterpart using some biochemical analysis along with inter simple sequence repeat (ISSR) marker. Afr. J. Biotechnol. 10(34):6401-6410.

Saurabh B, Sneha S, Suvidya R, Pramod K, Shailesh B (2011). Analysis of distribution and significance of simple sequence repeats in enteric bacteria Shigella dysenteriae SD197. Bioinformation 6(9):348-351.

Shikano T, Ramadevi J, Shimada Y, Merilä J (2010). Utility of sequenced genomes for microsatellite marker development in non-model organisms: a case study of functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). BMC Genomics 11: 334.

Singh A, Feschotte C, Stojanovic N (2007). Micro-repetitive structure of genomic sequences and the identification of ancient repeat elements, November 2-4, 2007 Proceedings (IEEE Int Conf Bioinformatics Biomed) pp.165-171.

Smit AFA, Green P (2002). RepeatMasker. unpublished. Website http://ftp.genome.washington.edu/RM/RepeatMasker.html.

Sun H, Mo Q, Lin J, Yang Z, Tu C, Gu D, Shi L, Lu W (2011). Rapid simultaneous screening of seven clinically important enteric pathogens using a magnetic bead based DNA microarray. World J. Microbiol. Biotechnol. 27(1):163-169.

Tarailo-Graovac M, Chen N (2009). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Curr. Protoc. Bioinformatics (Suppl.25).

Treangen TJ, Abraham A, Touchon M, Rocha EPC (2009). Genesis, effects and fates of repeats in prokaryoticgenomes. FEMS Microbiol. Rev. 1-33

Tyagi S, Sharma M, Das A (2011). Comparative genomic analysis of simple sequence repeats in three Plasmodium species. Parasitol. Res. 108(2):451-458.

van Belkum A (1999a). Short sequence repeats in microbial pathogenesis and evolution. Cell. Mol. Life Sci. 30:56(9-10):729-34.

van Belkum A (1999b). The role of short sequence repeats in epidemiologic typing. Curr. Opin. Microbiol. 2(3):306-311.

van Belkum A, Scherer S, van Alphen L, Verbrugh H (1998). Short-Sequence DNA Repeats in Prokaryotic Genomes. Microbiol. Mol. Biol. Rev. 62(2): 275-293.

van Belkum A, van Leeuwen W, Scherer S, Verbrugh H (1999). Occurrence and structure-function relationship of pentameric short sequence repeats in microbial genomes. Res. Microbiol. 150(9-10):617-626.

Volfovsky N, Haas BJ, Salzberg SL (2001). A clustering method for repeat analysis in DNA sequences. Genome Biol. 2(8):RESEARCH0027.

Von Graevenitz A, Weinstein J (1971). Pathogenic significance of Pseudomonas fluorescens and Pseudomonas putida. Yale J. Biol. Med. 44(3):265-273.

Wang L, Zhang A (2005). BioStar models of clinical and genomic data for biomedical data warehouse design. Int. J. Bioinformatics Res. Appl. 1(1):63-80.

Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W, Fournier G, Mayhew GF, G. Plunkett III, Rose DJ, Darling A, Mau B, Perna NT, Payne M, Runyen-Janecky LJ, Zhou S, Schwartz DC, Blattner FR (2003). Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T. Infect. Immun. 71(5):2775-2786.

Whatmore AM (2009). Current understanding of the genetic diversity of Brucella, an expanding genus of zoonotic pathogens. Infect. Genet. Evol. 9(6):1168-1184.

Yang J, Wang J, Chen L, Yu J, Dong J, Yao ZJ, Shen Y, Jin Q, Chen R (2003). Identification and characterization of simple sequence repeats in the genomes of Shigella species. Gene 322: 85-92.

Zhang J (2003). Evolution by gene duplication: an update. Trends Ecol. Evol. 18(6):292-298.

Full Length Research Paper

Sensitive and rapid detection of *Mycoplasma capricolum* subsp. *capripneumoniae* by Loop-mediated isothermal amplification

HE Ying[1], ZHANG Nian-zhang[1], ZHAO Ping, CHU Yue-feng, GAO Peng-cheng, ZHANG Jian-jun, LIU Xiang-tao and Lu Zhong-xin*

State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Epizootic Diseases of Grazing Animal of Ministry of Agriculture, Lanzhou 730046, China.

A set of four specific primers was designed by targeting the *H2* gene sequences of *Mycoplasma capricolum* subsp. *capripneumoniae* (MCCP). Using *Bst* DNA polymerase, the products were amplified for 60 min at 65°C in a simple water bath. Compared with a polymerase chain reaction (PCR) test that targets the *H2* gene sequences of MCCP, the sensitivity of the loop-mediated isothermal amplification (LAMP) assay was higher (approximately 0.75 fg DNA per reaction). The LAMP products could be visualized by agar gel electrophoresis. There were no cross reactions with other strains in the *Mycoplasma mycoides* cluster, which indicates the high specificity of the LAMP procedure. The LAMP assay was able to detect MCCP in tissue.

Key words: *Mycoplasma capricolum* subsp. *Capripneumoniae*, loop-mediated isothermal amplification, rapid detection.

INTRODUCTION

Contagious caprine pleuropneumonia (CCPP) is a severe infectious disease of goats caused by *Mycoplasma capricolum* subsp. *capripneumoniae* (MCCP), which occurs in many countries of Africa and Asia (Woubit et al., 2004). It is a disease of major economic relevance characterized by high morbidity and mortality; the mortality rate often approaches 100% in susceptible flocks (Rurangirwa et al., 1987). A mycoplasma strain, designated F-38, first isolated in Kenya (MacOwan and Minette, 1976), is a member of the *Mycoplasma mycoides* cluster which includes *M. mycoides* subsp. *mycoides* SC (MmmSC), *M. mycoides* subsp. *mycoides* LC (MmmLC), *M. mycoides* subsp. *capri* (mmc), *M. capricolum* subsp. *capricolum* (Mcca) and *Mycoplasma* species bovine group7 (bg7).

CCPP is classified as a list B disease by the Office International Des Epizooties or World Organization for Animal Health (OIE). It is an infectious disease that

Abbreviations: CCPP, Contagious caprine pleuropneumonia; **MCCP,** *Mycoplasma capricolum* subsp. *capripneumoniae*; **LAMP,** loop-mediated isothermal amplification; **PCR,** polymerase chain reaction.

affects only goats, and was first described in the late 19th century (Hutcheon, 1889; McMartin et al., 1980). Before the isolation and identification of *Mycoplasma* strain F38 by MacOwan (1976) and the subsequent demonstration of its causal relationship with CCPP (MacOwan and Minette, 1976), *M. mycoides* subsp. *capri* was considered to be the aetiological agent of CCPP (Edward, 1953; JonAs and Barber, 1969). So far, *M. capricolum* subsp. *capripneumoniae* is the only mycoplasma that fulfills the Koch postulates for CCPP, and it is believed to be the sole cause of CCPP (MacOwan, 1984). *Mycoplasma* strain F38 has been reclassified recently, and now all F38-like mycoplasmas are known as *Mycoplasma capricolum* subsp. *Capripneumoniae* (Leach et al., 1993).

CCPP has been reported to affect only goats (Thiaucourt and Bolske, 1996) and it does not cause disease in sheep, either spontaneously or experimentally (McMartin et al., 1980). However, there are some reports that describe the isolation of *M. capricolum* subsp. *capripneumoniae* from healthy sheep in Kenya that had been in contact with goat herds affected by CCPP (Litamoi et al., 1990), and from sick sheep in Uganda that had been mixed with goats suffering from the disease (Bolske et al., 1995). The isolation of *M. capricolum* subsp. *capripneumoniae* from cattle with mastitis has also been reported (Kumar and Garg, 1991), and these reports contradict the perceived host specificity of *M. capricolum* subsp. *capripneumoniae*.

Recently, a novel nucleic acid amplification method, termed loop-mediated isothermal amplification (LAMP), has been reported (MacOwan and Minette, 1976). The LAMP assay is rapid and its amplification efficiency is equivalent to that of polymerase chain reaction (PCR)-based methods (Cai et al., 2010; Gadkar and Rillig, 2008). More importantly, the approach is less costly, and all reactions can be developed in an isothermal environment. Reports of the detection of MCCP using LAMP assays have shown that the approach is easier and faster to perform than conventional PCR assays, as well as being more specific (Endo et al., 2004). In this study, a method based on the LAMP assay for the detection of CCPP was developed, and the sensitivity and specificity of the assay were evaluated. The assay was compared with a PCR test targeting the *H2* gene sequences of MCCP.

MATERIALS AND METHODS

Strains and cultivation

The origin of the 21 stains used in this study, type strains of the *M. mycoides* cluster and field isolates from China, are listed in Table 1. The mycoplasmas were cultivated in modified KM$_2$ (Hanks solution with 1.7% lactalbumin hydrolysate, 1% MEM, 20% de-complemented horse serum, 5% fresh yeast extract, 1% thallium acetate, 0.4% sodium pyruvate) in a high security laboratory. The DNA of *Pasteurella multocida* and *Mannheimia haemolytica* was maintained at the State Key Laboratory of Veterinary Etiological Biology.

Clinical samples

Twenty-eight (28) samples from 14 goats infected artificially with *M. capricolum* subsp. *capripneumoniae* were used in the study. The clinical samples were collected when the goats showed primary clinical signs: Cough, anorexia, laboured breathing with painful grunting, and a rise in temperature up to 41°C. Given that the gross pathological lesions were localized exclusively in the lung, the livers were collected for use in the experiment (Table 1). Sixty-one (61) clinical samples collected from western China in 2009-2011 were used for the epidemiological survey. These samples were kept at -80°C until analysis.

Sample preparation

Culture samples (1 ml) were centrifuged at 1,2000 RCF for 20 min at 4°C. The cell debris was pelleted; the pellets were washed in phosphate buffered saline (PBS) and re-suspended in 50 μl ddH$_2$O. After vortexing, the samples were lysed by boiling for 10 min, centrifuged and diluted 1:50. For the clinical samples, DNA extraction was performed using a kit (DNA extraction kit, Invitrogen, Carlsbad, USA) according to the protocol of the manufacturer. The extracted DNA was used for PCR both undiluted and at 1:50 dilution.

PCR conditions (He et al., 2011)

Based on the conserved sequence of the H2 gene (GenBank access number:AF162991.1) of MCCP, suitable primers were designed using the primer5.0 software: mccp f 5′ AAA AGT CCC TGA AAC ATT AC 3′ (319-338 bp) and mccp R5′ GGT GTA CCC ACT GCT AAA GA 3′ (1032-1013 bp). These primers were synthesized by TaKaRa, Dalian, China. The 50 μl reaction mixture contained 3 μl MgCl$_2$ (1.5 mM), 0.5 μl dNTP (150 μM for dCTP and dGTP, 300 μM for dATP and dTTP), 5 μl 10×Taq Buffer[b], 1 μl each primer, 0.5 μl Taq polymerase (10×Taq dNTP Buffer and Taq polymerase (1 unit, TaKaRa, Dalian, China)), 5 μl DNA sample and 34 μl ddH$_2$O. The PCR conditions consisted of an initial denaturation step of 5 min at 94°C, followed by 35 cycles of 45 s at 94°C, 30 s at 50°C and 45 s at 72°C, and a final extension step of 10 min at 72°C. Samples of the PCR amplification products (5 μl) were subjected to electrophoresis in a 1% agarose gel in tris/borate buffer according to standard protocols. The DNA was visualized by UV-fluorescence after staining with ethidium bromide. Each PCR test was repeated three times.

LAMP primers

The MCCP LAMP primer set was designed using the primerExplorer program http//primerexplorer.jp/e/index.html to amplify the H2 gene (GenBank access number: AF162991.1). The primers are shown in Table 2.

LAMP method

The LAMP reactions were carried out with a 25 μL reaction mixture containing 2 μL of extracted DNA, 40 pmol (each) of primers FIP and BIP, 5 pmol (each) of primers F3 and B3, 2.8 mM of each dNTP, 4 U of the large fragment of *Bst* DNA polymerase (*Bst* DNA polymerase (Biolabs[@]inc., New England, USA)), with the corresponding polymerase buffer. The reaction temperature and time were 65°C and 60 min. The reaction was terminated by heating at 80°C for 3 min. Positive and negative controls were

Table 1. Isolation of *Mycoplasma capricolum subsp.capripneumoniae* and LAMP and PCR-detection from artificial infected animals with MCCP.

Goat number	Macroscopic findings	Sample type	Microbiology results	LAMP 1:50 diluted	PCR 1:50 diluted
1	Serious hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
2	Slight hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
3	Slight hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
4	Serious hepatized	Lung	MCCP	+	+
5	Serious hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
6	Slight hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
7	Slight hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
8	Slight hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
9	Serious hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
10	Serious hepatized	Lung	-	+	+
		Liquor pleurae	-	+	-
11	Serious hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
12	Serious hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
		Liquor pleurae	MCCP	+	+
13	Serious hepatized	Lung	MCCP	+	+
		Liquor pleurae	MCCP	+	+
14	Serious hepatized	Lung	MCCP	+	+
	Hepatized	Liquor pleurae	MCCP	+	+

Table 2. Primer sequences used in this study.

Name	Sequence(5`-3`)
FIP	TGCTGGTGAATATTTTGTAGCAGGTTTTTAAGCCCAAAGTTAATATACCTTGA
BIP	CAACACCAGATTCAAAGAAAGGTTTTTTTTGAGTTGAAAGCTTTTTAGATTGT
F3	ACAACCTAAAGAGATTATTCACTC
B3	AACTTGACTTCCAACAACAA

included in each run, and all precautions to prevent cross contamination were observed. The LAMP products (3 µL) were detected in 2% agarose.

Specificity of LAMP

To determine the specificity of the LAMP method, it was carried out at 65°C for 60 min with the DNA of various mycoplasmas in the *M. mycoides* cluster; the templates from the eight type strains of the *M. mycoides* cluster and *Pasteurella multocida* are listed in Table 3. Each DNA sample from the strains tested was examined in

triplicate. The products were separated by 2% agarose gel electrophoresis, and the target bands were visualized by staining with ethidium bromide.

**Sensitivity of the LAMP method**

The sensitivity of the assay was assessed by testing tenfold serial dilutions of 1 µg ml^{-1} DNA of *M. capricolum* subsp. *capripneumoniae* F1601. Reaction mix without the DNA template was included as a negative reaction control. The LAMP amplification products were analyzed visually by 2% agarose gel electrophoresis. To compare

Table 3. Collection of strains used to test the specificity of the *Mycoplasma capricolum subsp.capripneumoniae* LAMP.

Species	N	Description	Origin	PCR
M. capricolum subsp. capripeumoniae	1	M1601	M1601, China	+
M. ovipneumoniae	1	Y98	Y98, ntcc	-
M. mycoides subsp. Capri	1	PG3	PG3, China	-
M. mycoides subsp. mycoides Large colony	1	Y-goat	Cirad	-
M. capricolum subsp. capricolum	1	C.Kid	Cirad	-
M. agalactiae	1	GS. 12	China	-
M. arginini	1	PG 1	China	-
M. bovis	1	M.B1	China	-
Pasteurella multocida	1	*Pasteurella multocida*	China	-

Table 4. Isolation of *Mycoplasma capricolum subsp.capripneumoniae* and LAMP and PCR-detection from clinical samples obtained from animals suspected for being infected with MCCP.

Sample type	Number of positive results (%)	
	LAMP	PCR
Lung tissue	33 (42)	26 (42)
Liquor pleurae	14 (19)	10 (19)
Total	47	36
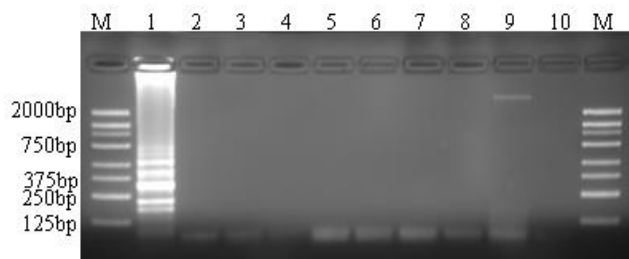
Figure 1. Gel dielectrophoresis of LAMP products showing a specific for *M. capricolum subsp. Capripeumoniae* (MCCP). Lane: M, Molecular weight; lane 1, M1601; lane 2, *M. mycoides susp. Capri* PG3; lane 3, *M. ovipneumoniae* strain Y98; lane 4, *Pasteurella multocida*; lane 5, *M. capricolum subsp. Capicolum* C.kid; lane 6, *M. agalactiae* GS.12; lane 7, *M. arginine* PG 1; lane 8, *M. bovis* strains M1; lane 9, *M. mycoides susp. mycoides large colony* Y-goat; lane 10, ddH2O.

the detection sensitivities of LAMP and PCR, PCR using *H2* gene primers that amplify a 680-bp product was carried out in a total reaction volume of 25 µL containing 2 µL of the MCCP DNA, 2 µL (10 pmol ml^{-1}) of a pair of appropriate primers, 12.5 µL Premix Ex Taq (TaKaRa, Shiga, Japan) and ddH$_2$O. The PCR conditions consisted of an initial denaturation step of 5 min at 94°C, followed by 35 cycles of 45 s at 94°C, 30 s at 50°C and 45 s at 72°C, and a final extension step of 10 min at 72°C. Samples of the PCR amplification products (5 µl) were analyzed by 1% agarose gel electrophoresis. Tests of the sensitivity of the LAMP and PCR assays were conducted in triplicate, and the detection limits were defined as the last positive dilutions, with a sample being considered positive if all three samples tested positive.

Results

Specificity of the LAMP assay

The specificity of LAMP was tested using DNA extracted from the eight type strains of the *M. mycoides* cluster and *P. multocida*. After incubation at 65°C for 60 min, MCCP was positively detected, whereas no other mycoplasma isolate was detected or amplified by LAMP. The other type strains of the *M. mycoides* cluster and *P. multocida*, as listed in Table 3, were negative (Figure 1).

Sensitivity of the LAMP assay

To assess the sensitivity of the LAMP assay for the detection of MCCP, the reaction was tested using 1 µL tenfold serial dilutions of MCCP DNA and compared with the PCR assay. The LAMP reaction was able to detect up to 0.75 fg DNA per reaction (Figure 2); however, the PCR could only detect MCCP up to 750 pg per reaction (Figure 3). The results indicate that LAMP has a higher sensitivity than the standard PCR method.

Clinical samples

Twenty-eight (28) clinical samples originating from 14 animals showing serious clinical signs were included in the analysis: 28 samples were positive by LAMP assay, using
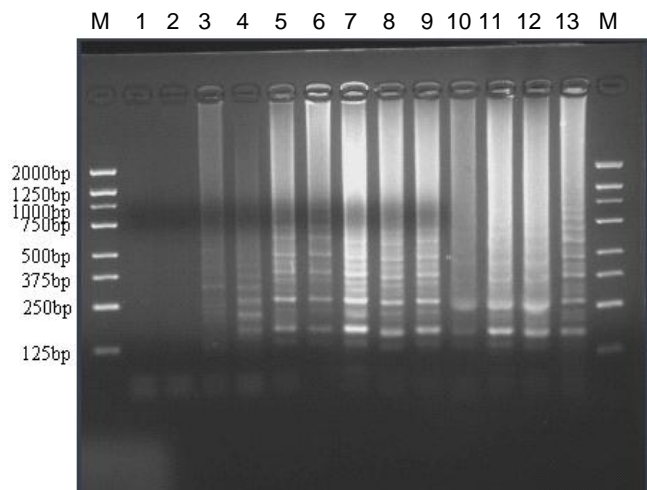
Figure 2. Sensitivity of LAMP. Lane: 13-3, 7.5 µg, 750 ng, 75 ng, 7.5 ng, 750 pg, 7.5 pg, 0.075 pg, 0.0075 pg, 0.00075 pg. The last concentration y detection was 0.00075 pg.
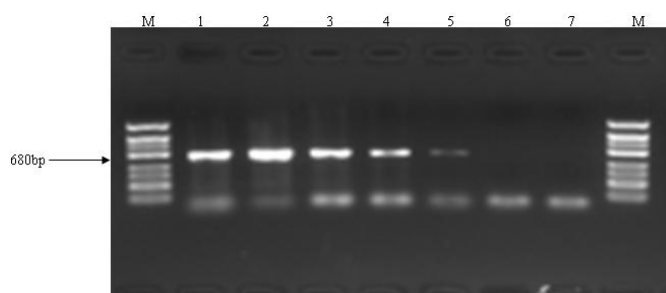

Figure 3. Sensitivity of PCR. Lanes: 1-5, represent 7.5 µg, 750 ng, 75 ng, 7.5 ng, 750 pg. The last concentration by detection was 750 pg.

1:50 diluted templates (Figure 4). To evaluate the LAMP test further, comparison with the PCR test was performed using samples from an epidemiological survey conducted in western China. The results are shown in Table 4. Overall, from 61 clinical samples, the LAMP assay gave a total of 11 more positive results than the PCR test.

DISCUSSION

The LAMP primers used in this study were based on the H2 gene sequence, which confers some advantages for molecular identification. The H2 gene is a putative membrane protein gene. It is reported that the partial sequence of the *H2* gene can be used as an epidemiological marker for the *M. mycoides* cluster because it is a conserved sequence. Such studies demonstrate the use of the *H2* gene for the molecular identification of closely related genomic species (Thiaucourt and Bolske,
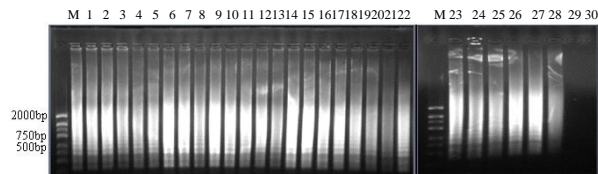

Figure 4. LAMP from artificial infected animals with MCCP and the samples was diluted as 1:50. Lane: M, Molecular weight marker; lanes 1-28, clinical samples; lane 29, negative control; lane 30, normal lung tissue of goat.

1996).

LAMP method compared with conventional PCR reported in this article has the advantages of simple operation, rapid reaction and ease of detection. The LAMP assay is a simple detection tool in which the reaction is performed in a single tube by mixing the thermopol buffer, primers and *Bst* DNA polymerase, followed by incubation of the mixture at 65°C for 60 min. The LAMP reaction is performed under isothermal conditions and it does not require expensive equipment: the only equipment needed for the LAMP reaction is a regular laboratory water bath or a heating block that can provide a constant temperature of 65°C. Moreover, the amplification efficiency is extremely high, there is no time required for thermal cycling, and inhibition reactions at later stages are less likely to occur than in standard PCR. In addition, LAMP amplifies DNA to higher concentrations than PCR, allowing convenient visualization of the products after the addition of SYBR Green I without gel electrophoresis. Hence, the LAMP assay could be developed into a field test.

In this study, the LAMP method for the detection of MCCP was found to be highly sensitive, because it could detect MCCP at 0.75 fg DNA per reaction, whereas by PCR, the detection of MCCP was possible only up to750 pg DNA per reaction. This indicates that the sensitivity of LAMP is higher than that of the standard PCR. The increased sensitivity may make LAMP a better choice than PCR for the detection of MCCP from cases of in apparent infection.

In conclusion, the LAMP method described in this study represents a new, sensitive, specific, and rapid protocol for the detection of MCCP. It may be applied in epidemiological surveys of contagious caprine pleuropneumonia.

ACKNOWLEDGEMENTS

REFERENCES

Bolske G, Johansson KE, Heinonen R (1995). contagious caprine pleuropneumonia in uganda and isolation of mycoplasma-capricolum subspecies capripneumoniae from goats and sheep. Vet. Record 137:594-594.

Cai SH, Lu YS, Wu ZH (2010). Loop-mediated isothermal amplification method for rapid detection of Vibrio alginolyticus, the causative agent of vibriosis in mariculture fish. Letters in Appl. Microbiol. 50:480-485.

Edward DGf (1953). Organisms of the pleuropneumonia group causing disease in goats. Veterinary Record 65:873-874.

Endo S, Komori T, Ricci G (2004). Detection of gp43 of Paracoccidioides brasiliensis by the loop-mediated isothermal amplification (LAMP) method. Fems Microbiology Letters 234:93-97.

Gadkar V, Rillig MC (2008). Evaluation of loop-mediated isothermal amplification (LAMP) to rapidly detect arbuscularmycorrhizal fungi. Soil Biol. Biochem. 40:540-543.

He Y, Zhang N, Zhao P (2011). Improvement of specific polymerase chain reaction(PCR) for the identification of *Mycoplasma capricolum* subsp. *capripneumoniae* .Afri. J. Biotechnol. 10:17291-17296.

Hutcheon D (1889). Contagious pleuro-pneumonia in goats at Cape Colony, South Africa. Vet. J. 29:399-404.

JonAs AM, Barber TL (1969). Mycoplasma mycoides var. capri isolated from a goat in Connecticut. J. Infect. Dis. 119:126-131.

Kumar A, Garg DN (1991). Isolation of mycoplasma F-38 from the milk of mastitic cows. Vet. Record 128:429.

Leach RH, Erno H, Macowan KJ (1993). Proposal for designation of F38-type caprine mycoplasmas as Mycoplasma capricolum ssp. capripneumoniae, new subspecies and consequent obligatory relegation of strains currently classified as Mycoplasma capricolum (Tully, Barile, Edward, Theodore, and Erno 1974) to an additional new subspecies, Mycoplasma capricolum ssp. capricolum, new subspecies. Intl. J. Syst. Bacteriol. 43:603-605.

Litamoi JK, Wanyangu SW, Simam PK (1990). Isolation of mycoplasma biotype f38 from sheep in Kenya. Trop. Anim. Health Productn. 22:260-262.

MacOwan KJ (1984). Role of mycoplasma strain F38 in contagious caprine pleuropneumonia. Israel J. M. Sci. 20:979-981.

MacOwan KJ, Minette JE (1976). A mycoplasma from acute contagious caprine pleuropneumonia in Kenya. Trop. Ani. Health Product. 8:91-95.

McMartin DA, MacOwan KJ, Swift LL (1980). A century of classical contagious caprine pleuropneumonia: from original description to aetiology. Brit. Vet. J. 136:507-515.

Rurangirwa FR, McGuire TC, Magnuson NS (1987). Composition of a polysaccharide from mycoplasma (F-38) recognised by antibodies from goats with contagious pleuropneumonia. Res. Vet. Sci. 42:175-178.

Thiaucourt F, Bolske G (1996). Contagious caprine pleuropneumonia and other pulmonary mycoplasmoses of sheep and goats. Revue Scientifique Et Technique-Office International Des Epizooties 15:1397-1414.

Woubit S, Lorenzon S, Peyraud A (2004). A specific PCR for the identification of Mycoplasma capricolum subsp capripneumoniae, the causative agent of contagious caprine pleuropneumonia (CCPP). Vet. Microbiol. 104:125-132.

Full Length Research Paper

Genetic diversity in Nigerian brinjal eggplant (*Solanum melongena* L.) as revealed by random amplified polymorphic DNA (RAPD) markers

Sifau, Mutiu Oyekunle[1,3]*, Akinpelu, Adejoke[1], Ogunkanmi, Liasu Adebayo[1], Adekoya, Khalid Olajide[1], Oboh, Bola Olufunmilayo[1] and Ogundipe, Oluwatoyin Temitayo[2]

[1]Department of Cell Biology and Genetics, University of Lagos, Lagos, Lagos State, Nigeria.
[2]Department of Botany, University of Lagos, Lagos, Lagos State, Nigeria.
[3]Molecular Biology Laboratory, Biotechnology Unit, National Centre for Genetic Resources and Biotechnology (NACGRAB), PMB 5382, Ibadan, Oyo State, Nigeria.

The taxonomy of *Solanum melongena* L., also known as brinjal eggplant, has remained difficult because previous studies to establish genetic relationships among taxa are mainly based on morphological features, which are insufficient to establish genetic affinities. In the present investigation, five highly polymorphic random amplified polymorphic DNA primers were used to describe the genetic similarity and diversity among its accessions in Nigeria. The results show a high level of polymorphism based on the banding patterns among the samples. This indicated a wide and diverse genetic base. Four distinct clusters were equally noticeable at a coefficient of 0.80 from the dendrogram generated. Generally, the grouping pattern clearly indicates that irrespective of variations in fruit traits among samples, they were still grouped together in different clusters with a high similarity coefficient value. This probably showed some genetic relatedness/closeness among the samples concerned. The study also revealed that there is no association between RAPD pattern and the geographic origin of accessions. These agreed with previously published data on the characterization of eggplant. The study disclosed that molecular methods coupled with morphological analyses could make proper classification of *S. melongena* and other *Solanum* species in Nigeria possible to achieve.

Key words: Eggplant, random amplified polymorphic DNA (RAPD), Data, polymorphism, taxonomy.

INTRODUCTION

Eggplant, *Solanum melongena* L., also known as aubergine or brinjal, is a member of the family Solanaceae. It is one of the few cultivated solanaceous species originating from the Old World and an important vegetable in central, southern and south-east Asia, and in a number of African countries (Kalloo, 1993). Brinjal eggplant belongs to the

very large genus *Solanum*, as well as its largest sub-genus, *Leptostemonum*, which includes many wild relatives, as well as other cultivated species, such as the Gboma eggplant (*Solanum macrocarpon* L.) and the scarlet eggplant (*Solanum aethiopicum* L.) grown mostly in Africa for their fruits and leaves. More than 200 *Solanum* species are known in Africa, with about 25 species indigenous in Nigeria (Gbile and Adesina, 1988; Burkill, 2000).

A large number of cultivars of eggplant are known and characterized by variation in morphology, physiology and biochemical features (Daunay et al., 2001a). Although India or Indo-China is considered to be the centre of eggplant diversity (Lester and Hassan, 1991; Behera et al., 2006), but the affinities of brinjal eggplant (*S. melongena*) to related species remain uncertain. The crossability and hybridization studies of *S. melongena* with its related species have been generally inconclusive and the results are often contradictory (Anis et al., 1994; Behera and Singh, 2002). Taxa that are morphologically similar to *S. melongena* are difficult to classify (Karihaloo and Gottlieb, 1995) and the relationship of the cultivated eggplant with the weedy form, *Solanum insanum* and its wild progenitor *Solanum incanum* is unclear (Lester and Hassan, 1990). The analysis of accessions derived from different geographical areas is important to study the genetic diversity. However, diverse geographic origin of two accessions cannot be considered as a parameter to describe actual genetic diversity (Skroch et al., 1998). The continuum of morphological variation, cross compatibility, and genetic distances which exist between advanced and primitive cultivars of eggplant, with weedy and wild forms and relatives provides a model system for the study of gene flow of traits affected by domestication between a crop and its spontaneous forms (Behera et al., 2006). Although an Old World taxon, *S. melongena* unexpectedly shares strong genetic similarities with New World *Solanum* species (tomato and potato).

The taxonomy of eggplant remains a challenge (Behera et al., 2006), because of species' large size, overlapping ecogeographical distribution (Levin et al., 2005), morphological plasticity, similar genomes (Okoli, 1988) and existence of swamps of natural hybrids (Oyelana and Ugborogho, 2008). The persistence of confusion in taxonomic classification of eggplant complex especially *S. melongena* is due to the fact that phylogenetic relationships established among taxa are mainly based on morphological features (Karihaloo and Rai, 1995), crossability (Hassan and Lester, 1990) and F_1 fertility (Lester and Hassan, 1991). These parameters are, however, insufficient to establish genetic affinities, because *S. melongena* can be crossed not only to putative progenitors but also to more distantly related species (Daunay et al., 1991). Moreover, because of the existence of a high level of morphological variability, morphological data can lead to ambiguous interpretations.

To overcome these problems, isozyme variation has been analysed (Lester and Hassan, 1991; Karihaloo and Gottlieb, 1995), interspecific crosses involving species of *Solanum* in Nigeria has been carried out (Oyelana and Ugborogho, 2008). Obute et al. (2006) also carried out cytogenetic studies on some brinjal eggplants of Nigerian origin. The genetic affinities at the DNA level have also been reported based on the analysis of chloroplast DNA (Sakata and Lester, 1997 ; Isshiki et al., 1998); random amplified polymorphic DNA (RAPD) (Karihaloo et al., 1995; Singh et al., 2006); amplified fragment length polymorphism (AFLP) (Mace et al., 1999; Furini and Wunder, 2004); simple sequence repeat (SSR) or sequence tagged microsatellite site (STMS) markers (Nunome et al., 2003; Behera et al., 2006) and inter-simple sequence repeat (ISSR) markers (Isshiki et al., 2008; Ali et al., 2013). But little has been done so far, to assess the genetic diversity within the Nigerian brinjal eggplant using molecular markers. The application of molecular markers offers new tools to complement morphological parameters to resolve taxonomic and phylogenetic difficulties associated with eggplant and its relatives. Resolution of these quagmires is important for conservation and genetic improvement of eggplant *Solanum* and relatives for the benefit of breeders and for its sustainable use. In the present investigation, RAPD markers were used to describe the genetic similarity as well as diversity among accessions of *S. melongena* collected from different locations in Nigeria. This was done to gain a better insight into the centre of diversity of this plant and locate the probable source in Nigeria.

MATERIALS AND METHODS

Sample collections and identification

Samples were collected from different locations within the North and South-West of Nigeria. While on the field, each sample collected was given a code for temporary identification pending the determination of their proper names. The collected samples were first compared with photographs, drawings and illustrations from existing sample collections. Details of the samples were obtained from the database using Flora of West Africa from where the specimen was collected in order to determine and authenticate the collected samples. This was achieved by the use of keys in the Flora. Voucher specimen, that is, Herbarium specimen were then prepared following the method of Ogundipe et al. (2009) and taken to expert for proper determination and identification.

Isolation of total genomic DNA

DNA was extracted from the fruit mesocarp of samples using the modified protocol of Dellaporta as described by Dellaporta et al. (1983). DNA concentration was estimated using Eppendorf BioPhotometer spectrophotometer (Eppendorf AG22331, Hamburg, Germany). The quality of the extracted DNA was then verified by electrophoresis on a 1% Agarose gel for 1 h 30 min at 60 V.

Thereafter, the gel was viewed under the ultraviolet (UV) Gel

Figure 1. Samples of *S. melongena* collected for this study and their different shapes and colours. A and B Black purple colour of samples 001 to 011; C, green colour of samples 012 to 021; D, White colour of Samples 022 to 024.

Documentation and Analysis Systems (UVdoc, GA-9000/9010 Version 12, Uvitec, Cambridge, UK).

Polymerase chain reaction (PCR)

Initial screening was done with thirty RAPD primers (Operon Tech-nologies Inc., USA) using DNA from six accessions. Five Operon primers (P11, V04, Q07, U19 and Q03) that are highly polymorphic and gave scorable amplifications were selected and used in the analysis of all the 24 genotypes. Total reaction volume for DNA amplification was 10 μl containing 1.0 μl of 10x TAE buffer, 3 μl of 10 mg/μl sample DNA, 1.0 μl MgCl₂ (2.5 mM final concentration), 0.8 μl mixture of 10 mM DNTP (200 μM final concentration), 20 (5% Tween), 20 (Polyoxyethylene Sorbitan Monolaurate) with 20 ethylene oxide units, 3.6 μl of distilled water, and 0.6 μl Taq DNA polymerase (1 U final concentration). PCR was then run on the Techne TC- 412 thermal cycler (Model FTC41H2D, Barloworld Scientific Ltd, Staffordshire, UK), using the following temperature profile: Initial strand separation step of 3 mins at 94°C followed by 45 cycles each consisting of a denaturing step of 1 min at 94°C, annealing step of 1 min at 37°C and an extension step of 1 min at 72°C. The last cycle was followed by 5 min extension at 72°C. After amplification, PCR product was stored at 4°C till electrophoresis.

PCR products (amplicon) were mixed with 2.5 μl of 10X loading dye (0.25% bromophenol blue, 0.25% xylene cyanol and 40% sucrose, w/v) preparatory to electrophoresis and spun briefly in a microcentrifuge before loading into wells formed by teeth of combs in the agarose gel (Sambrook and Maniatis, 1989). PCR products were then resolved by electrophoresis at 2% agarose gel with constant power of 75 V for 2 h followed by staining with ethidium bromide (10 mg/ml) for 1 min and then placed in distilled water to remove excess stain. After the electrophoresis the gel was visua-lized and photographed in the Gel Documentation and Analysis Systems (UVdoc, GA-9000/9010 Version 12).

Data analysis

Only distinct, well-resolved and unambiguous bands were scored. The amplified fragments were scored as 1 (present) and 0 (absent). From this binary matrix, similarity matrices were computed using Sequential Hierarchical and Nested (SAHN) clustering option of the NTSYS-pc 2.02j software package (Rohlf, 1996). The software generated a dendrogram (Figure 1), which grouped the test lines using unweighted pair group method with mathematic average (UPGMA) on the basis of genetic similarity and Jaccard's coeffi-cient.

RESULTS

Thirteen fruits with purple skin colour were collected from the North (Kano State) while 11 fruits with green, white and purple-green skin colour were collected from the South-West (Ondo State) as shown in Figure 1. Table 1 shows lists of some of the qualitative morphological

Table 1. Some qualitative morphological characters observed in the 24 samples of *S. melongena* used in this study.

Fruit serial code	Fruit colour	Fruit size	Fruit shape
001	Slightly dark purple	Big	Oblong
002	Dark Purple	Slightly Big	Elongated and Bent
003	Dark Purple	Big	Base bigger than top
004	Dark Purple	Big	Flat with wide base
005	Very Dark Purple	Big	Dumb bell
006	Light Purple	Medium	Oblong with curved base
007	Lilac	Medium	Elongated
008	Light Purple	Small	Elongated with curved base
009	Light Purple	Small	Slightly Elongated
010	Very Light Purple	Small	Oval shape
011	Dark Purple	Very Small	Bent base bigger than top
012	Slightly Dark Green	Medium	Elongated and Slightly bent
013	Slightly Dark Green	Small	Elongated
014	Light Green	Small	Base bigger than top
015	Dark Green	Medium	Elongated but fleshy
016	Dark Green	Small	Top equal to base
017	Light Green	Small	Base bent bigger than top
018	Light Green	Small	Oval
019	Light Green	Very Small	Elongated
020	Dark Green	Small	Curved with bigger base
021	Light Green	Small	Base double size of top
022	White	Big	Elongated
023	White	Small	Oval
024	Green and Purple	Medium	Base equal in size to top

characters observed on the accessions.

The RAPD analysis carried out on all 24 accessions produced a large number of distinct fragments for each primer. A total of 44 bands amplified by five different oligonucleotide primers, were scored among the 24 accessions. Thirty two (32) of these bands were highly polymorphic with percentage polymorphism put at 72.7%. The numbers of amplification products obtained were in the range 8–11 with the primers OPQ-03, OPQ-07 and OPU-19 producing the minimum number of (8) bands each whereas primer OPP-11 produced the maximum number (11) of bands. Average of 8.8 bands was obtained per primer as shown in Table 2. Figure 2 shows the RAPD profile produced by OPP–11 for the 24 accessions.

Jaccard's similarity coefficient matrix was used to generate a dendrogram (Figure 3) using UPGMA clustering option of NTSYS-pc 2.02j software package (Rohlf, 1996). The scale of the dendrogram constructed from the data generated was between 0.74 and 0.94 with a mean value of 0.84 (Figure 3). Four clusters were distinguishable from the dendrogram at a truncated line of 80% (a co-efficient of 0.80). Cluster I consisted of 4 accessions (1, 5, 12, and 20). Cluster II, which happened to be the

largest, consisted of 13 accessions represented by samples 6, 7, 15, 18, 22, 8, 14, 16, 13, 9, 10, 17 and 21. Cluster III consisted of samples 11, 19, 23 and 24 while Cluster IV being the smallest was made up of samples 2, 4 and 3 (Figure 3). The dendrogram showed clusters of *S. melongena* with a very high level of similarity among them. Meanwhile, in Cluster I members are morphologically different yet they were found in the same sub-cluster at a similarity level of 81.5%. Cluster II also consisted of members with morphological differences yet with 81.0% similarity level. Cluster III has members that were dissimilar in terms of fruit colour, size and shape yet they were found in the same cluster. Cluster IV members have a very high similarity level of 84.0%. Looking at the members, they have similar fruit colour as opposed to the fruit colour in members of Clusters I, II and III.

DISCUSSION

The molecular techniques such as RAPD and amplified fragment length polymorphism (AFLP) have been found to be useful and robust tool for detecting genetic diversity and determining genetic relationships within and among a

Table 2. Selected operon primers and their sequence as well as characteristics of amplification products in samples analysed.

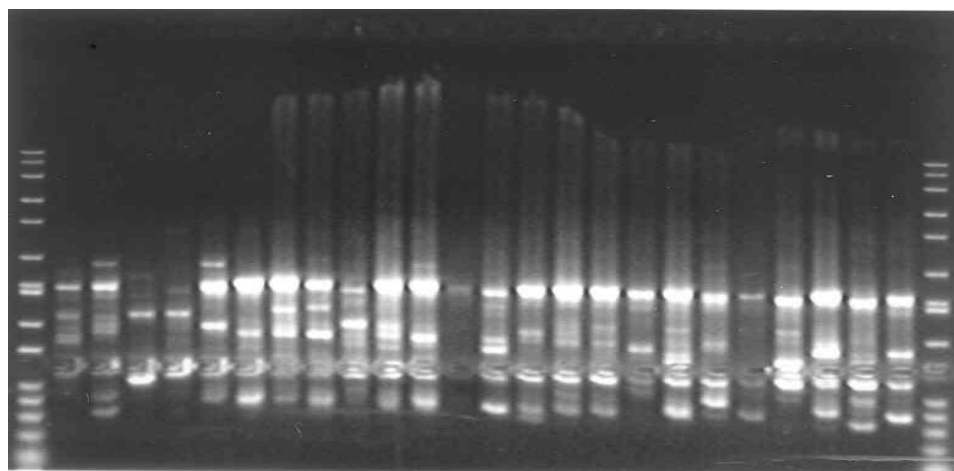| Primer used	Primer sequence (5´- 3´)	Number of Bands	Polymorphic bands	Percentage Polymorphic bands (%)
OPP – 11	(5´- AACGCGTCGG -3´)	11	9	81.8
OPV – 04	(5´- CCCCTCACGA -3´)	9	7	77.8
OPQ – 07	(5´- CCCCGATGGT -3´)	8	6	75.0
OPU-19	(5´- GTCAGTGCGG -3´)	8	5	62.5
OPQ – 03	(5´- GGTCACCTCA -3´)	8	5	62.5
Total		44	32	72.7
Average		8.8	6.4	

Figure 2. DNA profile produced by primer OPP-11 for samples studied. M represents the 100 bp DNA ladder which serves as the reference point; 1 to 24 corresponds to bands produced by the amplified DNA from the 24 *S. melongena* samples.

group of *Solanum* species (Mace et al., 1999; Singh et al., 2006; Toppino et al., 2008; Polignano et al., 2009; Ali et al., 2013). The present investigation, though focused on a limited number of accessions, gave useful information on the diversity of the *S. melongena*, their interrelationships and their importance in defining groupings characterized by different levels of similarity. High level of genetic diversity was observed in this study as earlier been reported among eggplant of the Indian subcontinent (Singh et al., 2006). Samuel (1996) had stated that high degree of diversity of species belonging to *Solanum* may be attributable to the fact that it is an ancient plant as well as its extraordinary rate of speciation. Singh et al. (2006) also stated that this high level of genetic variability in eggplant and related species was as a result of wide variation in the desirable genotypes/agronomy types in different regions.

The occurrence of accessions used in this study in different clusters despite the fact that they were deter-

mined to be the same species shows a genetic variability among them; hence the existence of large gene pool. Generally, the grouping pattern clearly indicates that irrespective of differences in fruit colour, shape and sizes that exist among the samples used, they were still interspersed with each other in different clusters (e.g. clusters I, II). This is an indication of some form of genetic relatedness/closeness among them despite differences in morphological features. The high value of similarity coefficient of 81.5 and 81.0% respectively is also another confirmation. This is in agreement with the findings of Singh et al. (2006).

A fairly high level of polymorphism was also observed in the present study with similarity coefficient ranging from 0.74 to 0.94, indicating a fairly wide and diverse genetic base. This observation is also in agreement with the results of some previous studies, such as that of Furini and Wunder (2004) and Polignano et al. (2009).

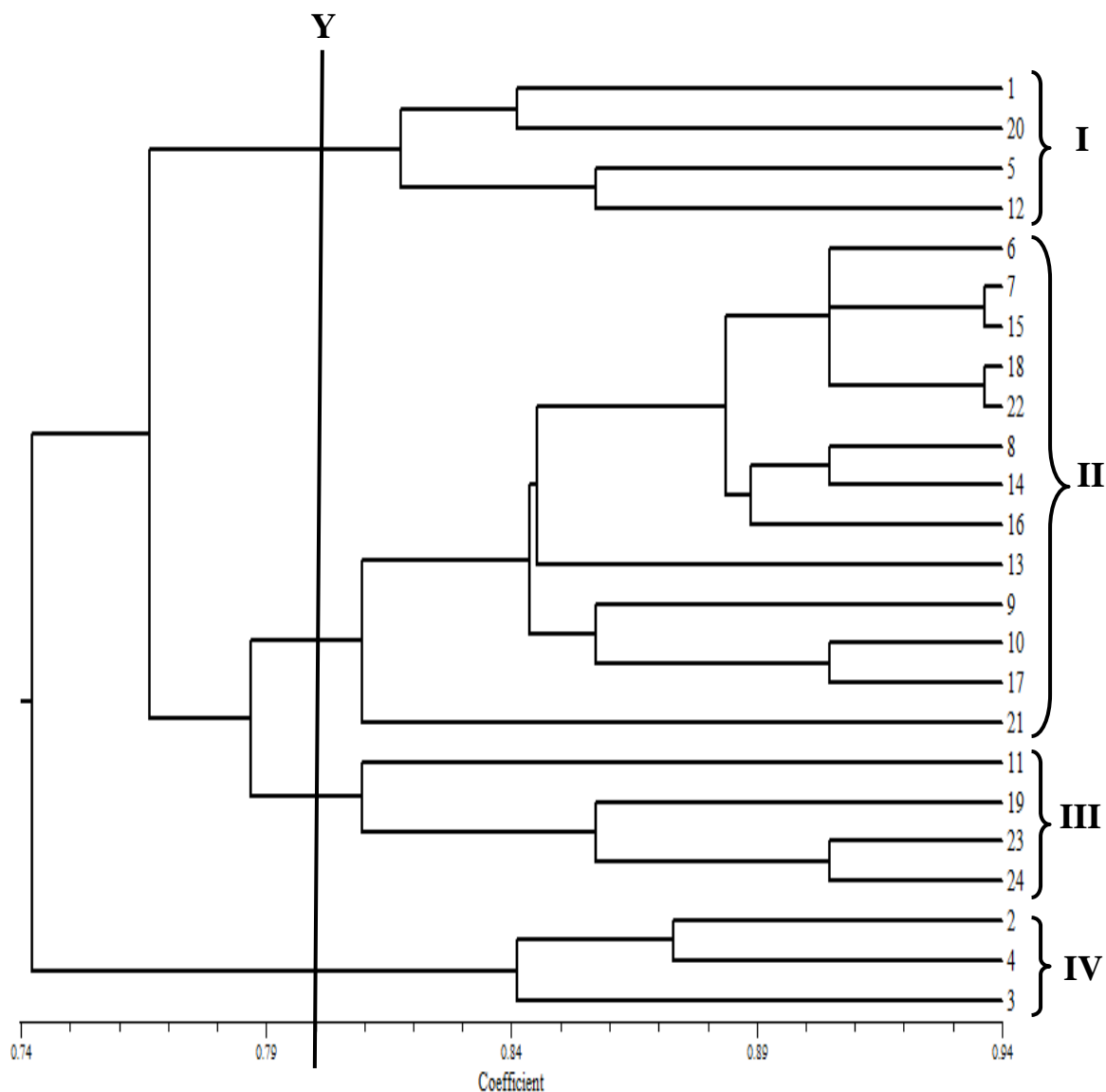The explanation for the observed high degree of variation

Figure 3. A dendrogram constructed by UPGMA cluster analysis from the RAPD data showing genetic relationship among accessions of *S. melongena* fruits. Y represents truncated line at a co-efficient of similarity 0.80; I to IV represents the four clusters that were distinguishable from the dendrogram at that co-efficient of similarity.

in this study could be due to the fact that the analysis of DNA variability is based on RAPD markers, which proved more informative than allozymes and morphological characters being used previously; and that the accessions of the *S. melongena* analysed in this study were collected from locations in Nigeria, where the greatest diversity has been reported (Gbile and Adesina, 1988). A high degree of variation has also been reported by using AFLP technology for *S. melongena* with weedy relative of the cultivated eggplant (Mace et al., 1999). However, this is a sharp contrast from the result of some other earlier workers who studied variation among the cultivated and weedy taxa of *S. melongena* by allozymes and RAPD

analysis. Examples include the work of Karihaloo et al. (1995) and that of Karihaloo and Gottlieb (1995). These authors observed little genetic polymorphism among the genotypes studied and came up with the suggestion that very small gene pool existed from which the cultivated forms arose.

Earlier workers on indigenous *S. melongena* in Nigeria based their characterization studies on cytogenetic (Obute et al., 2006), interspecific crosses (Oyelana and Ugborogho, 2008), morphological plasticity and similarity of genomes (Okoli, 1988). This present study happened to be the pioneer work on characterization and genetic variability of eggplant especially *S. melongena* in Nigeria

using molecular markers. Interestingly, collections originating from various parts of the country did not form well-defined distinct clusters as seen in Figure 3. They were interspersed with each other, indicating no association between RAPD pattern and the geographic origin of accessions. This finding is in agreement with the findings of Singh et al. (2006) and Polignano et al. (2009).

Conclusion

Although only a small sample of the gene pool was included in this study, within the eggplant *Solanum* the amount of variation may vary widely. Additional evaluations on the whole collection of *Solanum* species in Nigeria using different types of molecular markers (such as RAPD, AFLP, SNP, SSR, etc) as being done in other regions could give us more information on the genetic relationship among them. Detection of genetic differences and discrimination of genetic relationship between *Solanum* species could be a tool for utilization and conservation of plant genetic resources. *S. melongena* could also be a potential valuable breeding material use-ful to different users and assist in the introgression of genes. Nevertheless, the use of molecular methods cou-pled with morphological analyses could make proper classi-fication of different *Solanum* species in Nigeria possible to achieve. Finally, our results suggest further evaluation activities to better define the eggplant diversity patterns utilizing multivariate analysis and including larger sets from the Nigerian brinjal eggplant collection.

Conflict of Interests

The author(s) have not declared any conflict of interests

REFERENCES

Ali Z, Xu ZL, Zhang DY, He XL, Bahadur, Yi JX (2013). Molecular Diversity Analysis of Eggplant (*Solanum melongena*) Genetic Resources. Genet. Mol. Res. 10(2):1141-1155.

Anis M, Baksh S, Iqbal M (1994). Cytogenetic studies on F1 hybrid *S. incanum* x *S. melongena* var. American Wonder. Cytologia 59: 433–436.

Behera TK, Singh G (2002). Studies on resistance to shoot and fruit borer (*Leucinodes orbonalis*) and interspecific hybridization in eggplant. Ind. J. Hort. 59(1):62-66.

Behera TK, Sharma P, Singh BK, Kumar G, Kumar R, Mohapatra T, Singh NK (2006). Assessment of genetic diversity and species relationships in eggplant (*Solanum melongena* L.) using STMS markers. Sci. Hort. 107:352-357.

Burkill HM (2000). The Useful Plants of West Tropical Africa. 2nd Edition vol.5. Royal Botanical Gardens, Kew.

Daunay MC, Lester RN, Laterrot H (1991). The use of wild species for the genetic improvement of brinjal-eggplant (*Solanum melongena*) and tomato (*Lycopersicon esculentum*). In: J.G. Hawkes, R.N. Lester, M. Nee and N. Estrada-R (eds.), *Solanaceae III:* Taxonomy, Chemistry, Evolution. Royal Botanic Gardens, Kew, UK.

Daunay MC, Lester RN, Ano G (2001a). Cultivated eggplants in Tropical Plant Breeding. 2nd Edition, Oxford University Press, Oxford.

Dellaporta SL, Wood J, Hicks J. B (1983). A plant minipreparation: version II. Pl. Mol. Bio. Rep. 1:19-21.

Furini A, Wunder J (2004). Analysis of eggplant (*Solanum melongena*) related germplasm: morphological and AFLP data contribute to phylogenetic interpretations and germplasm utilization. Theor. Appl. Genet. 108:197-208.

Gbile ZO, Adesina SK (1988). Nigerian *Solanum* species of economic importance. Ann. Missouri Bot. Gard. 75(3):862-865.

Hassan SMZ, Lester RN (1990). Crossability relationships and *in vitro* germination of F1 hybrids between *Solanum melongena* L. X   *S. panduriforme* E. Meyer (*S. incanum* L. *sensu ampl*). SABRAO J. 22:65–72

Isshiki S, Uchiyama T, Tashiro Y, Miyazaki S (1998). RFLP analysis of a PCR amplified region of chloroplast DNA in eggplant and related *Solanum* species. Euphytica 102: 295-299.

Isshiki S, Iwataa N, Khana MMR (2008). ISSR variations in eggplant (*Solanum melongena* L.) and related *Solanum* species. Sci. Hort. 117:186-190.

Kalloo G (1993). Eggplant, *Solanum melongena* L. In: Genetic improve-ment of vegetable crops. Kalloo, G., Berg B.O. (Eds.). Pergamon Press, Oxford, UK.

Karihaloo JL, Gottlieb LD (1995). Allozyme variation in the eggplant, *Solanum melongena* L. (Solanaceae). Theor. Appl. Genet. 90:578–583.

Karihaloo JL, Rai M (1995). Significance of morphological variability in *Solanum insanum* L. *(sensu lato).* Plant Genet Res Newslett. 103:24–26.

Karihaloo JL, Brauner S, Gottlieb LD (1995). Random amplified polymorphic DNA variation in the eggplant, *Solanum melongena* L. (Solanaceae). Theor. Appl. Genet. 90:767–770.

Lester RN, Hassan SMZ (1990). The distinction between *Solanum incanum* L. and *Solanum insanum* L. (Solanaceae). Taxon 39: 521–523.

Lester RN, Hassan SMZ (1991). Origin and domestication of the brinjal eggplant, *Solanum melongena* from *S. incanum* in Africa and Asia. In *Solanaceae III:* Taxonomy, Chemistry, Evolution (eds. Hawkes, J.G., Lester, R.N., Nee, M. and Estrada, N.), The Royal Botanical Gardens, Kew, London.

Levin RA, Watson K, Bohs L (2005). A four-gene study of evolutionary relationship in *Solanum* Section Acanthophora. Amer. J. Bot. 92(4):603-612.

Mace ES, Lester RN, Gebhardt CG (1999). AFLP analysis of genetic relationships among the cultivated eggplant, *Solanum melongena* L., and wild relatives (*Solanaceae*). Theor. Appl. Genet. 99:626-633.

Nunome T, Suwabe K, Iketani H, Hirai M, Wrickle G (2003). Identi-fication and characterization of microsatellites in eggplant. Plant Breeding 122:256-262.

Obute GC, Benjamin CN, Okoli BE (2006). Cytogenetic studies on some Nigerian species of *Solanum* L. (Solanaceae). Afr. J. Biotechnol. 5(9):689-692.

Ogundipe OT, Ajayi GO, Adeyemi TO (2009). Phytoanatomical and Antimicrobial Studies on *Gomphrena celosioides* Mart. (Amarantha-ceae). Hamdard Medicus 51(3):146–156.

Okoli BE (1988). Cytotaxonomic studies of five West African species of *Solanum* L. (Solanaceae). Feddes Repert. 99(5-6):183-187.

Oyelana OA, Ugborogho RE (2008). Phenotypic variation of F1 and F2 populations from three species of *Solanum* L. (Solanaceae). Afr. J.

Biotechnol. 7(14):2359-2367.

Polignano G, Uggenti P, Bisignano V, Gatta CD (2009). Genetic divergence analysis in eggplant (*Solanum melongena* L.) and allied species. Genet. Res. Crop Evol. 57:171–181.

Rohlf FJ (1996). NTSYS-pc: numerical taxonomy and multivariate system, version 2.02j. Exeter Software. Setauket, New York.

Sakata Y, Lester RN (1997). Chloroplast DNA diversity in brinjal eggplant (*Solanum melongena* L.) and related species. Euphytica 97:295-301.

Sambrook JE, Fritch F, Maniatis T (1989). Molecular Cloning – A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, New York, USA. Sambrook JE, Fritch F, Maniatis T (1989). Molecular Cloning – A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, New York, USA.

Singh AK, Singh M, Singh AK, Singh R, Kumar S, Kalloo G (2006). Genetic diversity within the genus *Solanum* (Solanaceae) as revealed by RAPD markers. Curr. Sci. 90(5):711-716.

Skroch PW, Nienhuis J, Beebe S, Tohme J, Pedraza F (1998). Comparison of Mexican common bean (*Phaseolus vulgaris* L.) core and reserve germplasm collections. Crop Sci. 38:488–496.

Toppino L, Vale G, Rotino GL (2008). Inheritance of *Fusarium* wilt resistance introgressed from *Solanum aethiopicum* Gilo and Aculeatum groups into cultivated eggplant (*S. melongena*) and development of associated PCR-based markers. Mol. Breed 22(2):237-250.

Full Length Research Paper

Verification of genetic identity of introduced cacao germplasm in Ghana using single nucleotide polymorphism (SNP) markers

Jemmy Takrama[1], Ji Kun[2,3], Lyndel Meinhardt[2], Sue Mischke[2], Stephen Y. Opoku[1], Francis K. Padi[1] and Dapeng Zhang[2]*

[1]The Cocoa Research Institute of Ghana (CRIG), PO Box 8, New Tafo-Akim, Ghana.
[2]USDA/ARS, Beltsville Agricultural Research Center, SPCL, 10300 Baltimore Avenue, Bldg. 001, Rm. 223, BARC-W, Beltsville, MD 20705, USA.
[3]Southwest University, No.2 Tiansheng Road, Beibei, Chongqing, 400715, P.R. China.

Accurate identification of individual genotypes is important for cacao (*Theobroma cacao* L.) breeding, germplasm conservation and seed propagation. The development of single nucleotide polymorphism (SNP) markers in cacao offers an effective way to use a high-throughput genotyping system for cacao genotype verification. In the present study, high-throughput genotyping with SNP markers was used to fingerprint 160 cacao trees in the germplasm collection at the Cocoa Research Institute of Ghana (CRIG). These accessions had been originally introduced from international germplasm collections. The multilocus SNP profiles, generated by the Sequenom Mass Spectrometry platform, were compared with the SNP profiles of reference trees maintained in the international cacao collections. The comparison unambiguously identified mislabeled trees. For materials introduced as hybrid seeds without an available reference genotype, parentage analysis and model-based assignment were applied to verify their recorded parentage and genetic background. Our study shows that a small set of polymorphic SNP markers can provide a robust and accurate result for cacao genotype identification. This protocol can be applied for large-scale genotyping of cacao as well as for many other crops.

Key words: Cacao, conservation, chocolate, DNA fingerprint, molecular marker, tropical plant, off-type, true-to-type, West Africa.

INTRODUCTION

Cacao (*Theobroma cacao* L.) is an important tropical tree crop that provides raw ingredients for the chocolate confectionery industries. This global commodity has an annual production that exceeded 4 million tons in 2010,

Abbreviations: SNP, Single nucleotide polymorphism; **CRIG,** Cocoa Research Institute of Ghana; **EST,** express sequence tag; **PIC,** polymorphic information index.

of which 75% was produced in West Africa. Ghana alone produced 850,000 tons of cacao, accounting for 21% of the world's total output in 2010 (FAOSTAT, http://faostat3.fao.org/home/index.html). Cacao originated in the Amazon rainforest in South America and was domesticated by the Maya and Olmec peoples at least 3000 years ago (Cuatrecasas, 1964; Wood and Lass, 1985; Bartley, 2005; Powis et al., 2011). Beginning in the late 1800's and continuing into recent times, cacao has been repeatedly introduced into Ghana. Germplasm was ultimately deposited in an *in situ* germplasm bank at the Cocoa Research Institute of Ghana (CRIG) in Tafo, which currently houses over 1200 clones of various genetic origins (Edwin and Masters, 2005; Adu-Ampomah et al., 2006). Cacao is an outcrossing species (Wood and Lass, 1985) and germplasm is conserved as clonally propagated trees in field genebanks. Cacao germplasm collections have been shown to contain a variety of mislabeled individuals, and mislabeling is estimated at 15 to 44% in global cacao collections (Motilal and Butler, 2003; Motilal, 2004; Sounigo et al., 2006; Takrama et al., 2005). Mis-identifications can be attributed to multiplicity of introductions and transfers of plants from point-of-collection to establishment in early holding sites, and to subsequent recollection of budwood and repropagation of material for establishment. The potential for human error during plot demarcations and planting also contributes to this problem. Molecular markers have been used to characterize cacao germplasm since the 1980s (Guiltinan et al., 2008). Mislabeled accessions were identified by using dominant markers (Figueira et al., 1994; Whitkus et al., 1998; Sounigo et al., 2005) as well as codominant DNA markers such as restriction fragment length polymorphisms (Lerceteau et al., 1997; N'Goran et al., 2000). The development of microsatellite markers (Lanaud et al., 1999) greatly increased the efficiency and capacity for cacao fingerprinting and resulted in a wide application of cacao genotype identification (Aikpokpodion et al., 2005; Motilal and Butler, 2003; Efombagan et al., 2008; Motilal et al., 2010).

Recent progress in the development of cacao genomic resources has led to the use of single nucleotide polymorphisms (SNPs) as markers for cacao DNA fingerprinting, since SNPs are the most abundant class of polymorphisms in plant genomes (Buckler and Thornsberry, 2002). Compared with SSR markers, the assays of SNPs can be done without requiring separation of DNA by size, and therefore can be automated in an assay-plate format or on microchips. The diallelic nature of SNPs results in a much lower error rate in allele calling, and the genotyping can be multiplexed, allowing quicker completion at a lower cost than with SSRs. In recent years, SNP markers have been developed to assist cacao breeding and germplasm management (Allegre et al., 2012; Kuhn et al., 2012). TaqMan-based SNP assays have been developed for cacao genotyping under field conditions (Livingstone et al., 2012; Takrama et al., 2012). Using a set of SNP

markers derived from express sequence tag (EST) databases, Ji et al. (2013) characterized farmer selections of cacao from Nicaragua and Honduras and demonstrated that the SNP markers constitute a cost-effective marker resource suitable for cacao germplasm characterization. Results for genotyping with SNPs can be compared across different genotyping platforms and laboratories, facilitating the integration and interpretation of SNP data across different genebanks in various cacao-producing countries. The objective of the present study was to test the efficacy of using high-throughput SNP genotyping for molecular characterization of cacao and to assess the extent of mislabeling, or off-type, in the CRIG cacao germplasm collection.

MATERIALS AND METHODS

Sample preparation and SNP genotyping

One hundred and sixty (160) trees from the CRIG germplasm collection, representing 39 cacao accessions (each accession included one to five trees), were sampled for this experiment. Samples were collected from eight plots in the germplasm collection: D8 (2), L6 (34), M6 (32), M6 Ext. (5), Q6 (67), Q6 Ext.2 (8), Q6 Ext.4 (9), and V3 (3) (Table 1). Two young leaves were collected from each individual cacao tree and each sampled branch was tagged for potential revisiting. Both accession name and DNA extraction number were used to label each sample. DNA was extracted from the CRIG samples using the CTAB DNA Extraction Protocol (Doyle and Doyle, 1990). In addition, one hundred international clones were used as references. Preparation of DNA samples for the reference international clones was described in Zhang et al. (2009a; b). DNA concentration was determined with a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE). Based on the level of polymorphism and on their distribution across the ten chromosomes in cacao, 54 SNP markers were selected from 1560 candidate SNPs that had been developed using cDNA sequences from a wide range of cacao tissues (Argout et al., 2008). SNP genotyping was performed at the Human Genetics Division Genotyping Core facility, Washington University, St. Louis, using MALDI-TOF mass spectrometry (Sequenom, Inc., San Diego, CA). The heterozygosity and polymorphic information index (PIC) of these SNP markers has been reported by Ji et al. (2013).

Data analysis

Key descriptive statistics for measuring the informativeness of the SNP markers were calculated, including observed heterozygosity, expected heterozygosity, minor allele frequency, inbreeding coefficient and probability of identity (Evett and Weir, 1998; Waits et al., 2001). The program GenAIEx 6.2 (Peakall and Smouse, 2006; 2012) was used for computation. For the identification of mislabeling (off-types), SNP profiles of 100 reference trees maintained in the International Cacao Genebank, Trinidad (ICG,T) were used in the analysis. The genetic identity of the 100 reference trees has been characterized by both SNP (D. Zhang, USDA/ARS, Beltsville, personal communication) and SSR fingerprinting (Zhang et al., 2009b; Motilal et al., 2010; Johnson et al., 2009). Pairwise multi-locus matching was applied among each pair of individual trees, including the reference trees from the international germplasm collections, using the same program. Accessions with same names as the reference trees, but not matching them, were declared off-

Table 1. List of the 39 cacao accessions (represented by 160 trees), their field plot and tree stand, from Ghana cacao germplasm collection.

Sample code	Accession name	Number of trees	Plot number	Tree stand
1	ALPHAB 36	5	M6 ext	26; 9; 19; 28; 29
2	AMAZ 3-2	3	Q6 ext 2	225; 260; 243
3	CATONGO	1	Q6	1618
4	EQX 3338	3	Q6	1354; 1536; 1355
5	ICS 43	4	Q6 ext 4	10; 340; 12; 11
6	ICS 95	5	Q6	56; 143; 368; 429; 182
7	IMC 67	5	L6	364; 331; 250; 215; 331
8	IMC 76	5	Q6	738; 737; 729; 727; 724
9	MAN 15-60	5	Q6 ext 2	7; 13; 46; 85; 83
10	MOCORONGO	5	Q6	366; 367; 305; 428; 427
11	NA 33	5	M6	40; 22; 20; 19; 18
12	NA 79	2	Q6	959; 957
13	NA 79	3	Q6	960; 952; 958
14	P 30	5	Q6	906; 904; 901; 900; 898
15	PA 121	4	L6	102; 1396; 1282; 1453
16	PA 150	5	Q6	1019; 1013; 1012; 1010; 1009
17	PA 151	4	L6	687; 850; 686; 689
18	PA 300	5	L6	703; 737; 738; 702; 1246
19	PA 303	4	Q6	777; 1315; 1493; 892
20	PA 7	4	Q6	838; 839; 840; 844
21	PA 70	5	Q6	142; 205; 268; 403; 463
22	PA 88	4	Q6	299; 236; 360; 532
23	POUND 10	5	L6	1146; 843; 996; 995; 994
24	POUND 15	5	L6	1152; 1259; 780; 746; 1553
25	POUND 7	5	Q6	670; 672; 521; 673; 669
26	SCA 12	5	L6	254; 327; 253; 252; 211
27	SCA 6	5	Q6 ext 4	279; 278; 277; 276; 281
28	SCA 9	3	L6	194; 155; 154
29	SUL7	1	Q6	583
30	T16/613	2	D8	130; 128
31	T16/613	3	M6	131; 150; 151
32	T17/524	1	V3 1st planting	515
33	T60/877	5	M6	470; 473; 452; 450; 449
34	T63/967	5	M6	14; 15; 35; 17; 16
35	T63/971	5	M6	8; 9; 12; 10; 11
36	T65/238	5	Q6	1295; 1299; 1301; 1306; 1307
37	T65/326	2	V3 1st planting	745; 180
38	T79/501	4	M6	162; 143; 118; 121
39	T85/799	5	M6	250; 252; 231; 230; 229

types. For the multilocus matching, the option to ignore missing data was used. Discriminating power of the SNP loci was computed using the probability of identity (PID) (Waits et al., 2001) option implemented in the same computer program.

For accessions without a reference tree but with known pedigree record (for example, breeding lines selected in Ghana's breeding program), the genetic identities were verified using parentage analysis and/or model-based assignment test. An example is the T clones (Table 1) that were hybrid families introduced into West Africa in 1944. Since these were the products of hybridization in

Trinidad in the early 1940s, and the seed families were evaluated and selected in Ghana (Posnette, 1986), there are no existing reference trees available from the international cacao collections. Nonetheless, because pedigree records for these selections are available, the T clones were used as "offspring" and their parental clones in ICG,T were verified according to the recorded pedigree (Lockwood and Gyamfi, 1979). A likelihood-based method implemented in the program CERVUS 3.0 (Marshall et al., 1998; Kalinowski et al., 2007) was used for computation. For each parent-offspring pair, the natural logarithm of the likelihood ratio (LOD

score) was calculated.

Critical LOD scores were determined for the assignment of parentage to a group of individuals without knowing the maternity or paternity. Simulations were run for 10000 cycles with the assumption that 80% of candidate parents were sampled and a total of 80% of loci were typed, with a typing error rate of 0.5%. The most probable single mother (or father) for each offspring was identified on the basis of the critical difference in LOD scores (Δ) between the most likely and next most likely candidate parent at greater than 95% or 80% confidence (Marshall et al., 1998; Kalinowski et al., 2007).

For accessions lacking a reference tree, assignment test was applied to infer their hidden membership to a known population or germplasm group, using a model-based clustering analysis implemented in the STRUCTURE software program (Pritchard et al., 2000). SNP profiles of 100 reference accessions were included in the analysis. These 100 accessions were taken from six known Forastero germplasm groups, including Amelonado, Scavina (SCA) and Ucayali, Iquitos Mixed Calabacillo (IMC), Morona (MO), Nanay (NA) and Parinari (PA). Classification of these accessions have been reported by (Motamayor et al., 2008; Zhang et al., 2009b). The number of clusters (K-value, which indicated the number of sub-populations of the program attempted to find) was set from two to ten, and the analysis was carried out without assuming any prior information about the genetic group or geographic origin of the samples. Ten independent runs were assessed for each fixed number of clusters (K). The ΔK value was computed to detect the most probable number of clusters (Evanno et al., 2005). Of the 10 independent runs, the one with the highest Ln Pr (X|K) value (log probability or log likelihood) was chosen and represented as a bar plot.

RESULTS

Descriptive statistics of the SNP markers

In total, 53 SNP markers were reliably scored, as assessed by markers producing less than 10% missing genotypic data. Marker TcSNP 174 failed to generate SNP data thus was excluded in subsequent data analysis. The descriptive statistics of the remaining 53 SNP loci are presented in Table 2. The 53 SNP markers were polymorphic across the 39 cacao accessions. The mean expected heterozygosity was 0.343 and the observed heterozygosity was 0.274. An inbreeding coefficient with an average of 0.218 was observed.

Multilocus matching

Comparison of the multilocus SNP profiles with the reference accessions identified seven intraclonal mislabelings in accessions NA 79, PA 150 and IMC 76 (Figure 1). The multilocus matching also found that AMAZ 3-2 and PA 303 were mislabeled. These trees were defined as off-type or homonymous mislabeling because they shared the same name with the reference tree but differed in multilocus SNP profiles. In this experiment the mismatched accessions differed at a minimum of five loci. With all 53 loci considered, the combined probability of identity was in the order of 10^{-9} (Table 2). Overall, the procedure of multilocus matching with known reference trees led to the identification of 149 true-to-type trees out of

160 tested samples. Based on the verified result, 39 samples (a single sample from each accession) were used in the subsequent analyses of population structure and genealogical relationships. Among these 39 samples, the status of the nine T clones could not be decided solely based on multilocus matching, because they were selections made in Ghana and no reference trees were available. For these trees, assignment test and parentage analysis were applied to verify their genetic identity.

Assignment test

Based on the value of delta K, the model-based approach of STRUCTURE indicated K=5 as the most probable number of genetic clusters. The 39 tested cacao accessions from the Ghana cacao collection, as well as the 100 reference accessions, were stratified as germplasm groups of Amelonado, IMC, SCA/Ucayali, Morona, Nanay and Parinari, respectively (Figure 2). The assignment result largely agreed with the previously classified germplasm groups (Figure 2; Zhang et al., 2009b; Motamayor et al., 2008) except that the germplasm groups of SCA/Ucayali and Morona were not separated. The assigned memberships for all the tested trees from Ghana were compatible with their known parentage germplasm groups (Figure 2). The assignment test of the T clones confirmed their recorded parental germplasm groups, as shown in Figure 2. The parental groups of PA and IMC were clearly reflected in the admixed ancestry profiles of T60, T63, T65 and T79. A full genetic background of IMC was revealed for accession T85/799, supporting its recorded parentage of IMC 60 and NA 34 (a member of the IMC germplasm group; Motamayor et al., 2008). In addition, admixed ancestry of IMC and Amelonado was revealed for T16/613 family, which not only supported the recorded parentage of IMC 24, but also detected that the other parent came from the Amelonado group.

Parentage analysis

Of the eight candidate parent-offspring relationships, the results of parentage inference confirmed six pairs at the 95% confidence level and one pair (NA 34 - T85/799) at the 80% confidence level (Table 3). For offspring T16/613, only one parent (Amelonado 22) was identified at the >80% confidence level because the reference genotype of maternal parent IMC 24 was not available. The result of parent-offspring assignment supported the outcome of model-based clustering analysis by the STRUCTURE program (Figure 2).

DISCUSSION

Multilocus matching

Over 50 cacao germplasm collections are present worldwide

Table 2. Observed and expected heterozygosities, inbreeding coefficient, minor allele frequency and probability of identity of the 53 SNP loci scored on 39 cacao accessions from the Ghana Cacao germplasm collection.

Locus	Ho	He	Inbreeding coefficient	Minor allele frequency	PID-sib
TcSNP75	0.091	0.127	0.285	0.068	0.879
TcSNP90	0.091	0.127	0.285	0.068	0.879
TcSNP139	0.364	0.483	0.248	0.409	0.604
TcSNP144	0.523	0.500	-0.046	0.489	0.594
TcSNP150	0.310	0.436	0.290	0.321	0.635
TcSNP151	0.273	0.416	0.345	0.295	0.649
TcSNP189	0.295	0.425	0.305	0.307	0.642
TcSNP193	0.364	0.416	0.127	0.295	0.649
TcSNP226	0.318	0.397	0.198	0.273	0.662
TcSNP230	0.409	0.499	0.180	0.477	0.594
TcSNP242	0.488	0.447	-0.093	0.337	0.628
TcSNP309	0.500	0.375	-0.333	0.250	0.678
TcSNP329	0.295	0.312	0.052	0.193	0.725
TcSNP364	0.045	0.087	0.476	0.045	0.916
TcSNP372	0.182	0.201	0.097	0.114	0.814
TcSNP448	0.025	0.117	0.787	0.063	0.888
TcSNP469	0.227	0.298	0.236	0.182	0.736
TcSNP529	0.364	0.474	0.233	0.386	0.610
TcSNP534	0.477	0.487	0.021	0.420	0.602
TcSNP560	0.385	0.479	0.197	0.397	0.607
TcSNP577	0.366	0.442	0.172	0.329	0.631
TcSNP591	0.341	0.487	0.300	0.420	0.602
TcSNP602	0.341	0.456	0.253	0.352	0.622
TcSNP619	0.295	0.487	0.394	0.420	0.602
TcSNP702	0.295	0.469	0.370	0.375	0.614
TcSNP723	0.114	0.146	0.224	0.080	0.862
TcSNP731	0.278	0.424	0.345	0.306	0.643
TcSNP799	0.045	0.044	-0.023	0.023	0.956
TcSNP823	0.318	0.491	0.352	0.432	0.600
TcSNP872	0.317	0.495	0.360	0.451	0.597
TcSNP878	0.114	0.184	0.381	0.102	0.829
TcSNP886	0.227	0.416	0.453	0.295	0.649
TcSNP891	0.000	0.480	1.000	0.400	0.606
TcSNP899	1.000	0.500	-1.000	0.500	0.594
TcSNP928	0.024	0.024	-0.012	0.012	0.977
TcSNP998	0.231	0.416	0.445	0.295	0.649
TcSNP999	0.000	0.176	1.000	0.098	0.836
TcSNP1038	0.045	0.165	0.725	0.091	0.845
TcSNP1060	0.386	0.339	-0.141	0.216	0.704
TcSNP1063	0.091	0.127	0.285	0.068	0.879
TcSNP1111	0.114	0.146	0.224	0.080	0.862
TcSNP1126	0.045	0.087	0.476	0.045	0.916
TcSNP1159	0.068	0.066	-0.035	0.034	0.936
TcSNP1253	0.279	0.357	0.218	0.233	0.691
TcSNP1280	0.364	0.496	0.267	0.455	0.596
TcSNP1309	0.465	0.422	-0.103	0.302	0.645
TcSNP1331	0.326	0.487	0.331	0.419	0.602
TcSNP1378	0.455	0.351	-0.294	0.227	0.695
TcSNP1439	0.250	0.363	0.312	0.239	0.686
TcSNP1442	0.295	0.442	0.331	0.330	0.631

Table 2. Contd.

TcSNP1453	0.116	0.110	-0.062	0.058	0.895
TcSNP1458	0.386	0.479	0.194	0.398	0.607
TcSNP1484	0.523	0.479	-0.091	0.398	0.607
Mean	0.274	0.343	0.218	0.262	$6.5 \times 10^{-9*}$

* Accumulated PID_sibs for 53 SNP locus combinations.

Name	Field stand	Genotype	139	144	150	151	189	193	230	242	309	529	534	591	602	619	702	886	1060	1253	1280	1378	1484
PA 150	1019Q6	√	GG	AC	GG	CC	GG	AA	GG	CC	TT	AC	CT	AC	CT	TT	CC	CT	CC	TT	AA	TT	AG
PA 150	1013Q6	√	GG	AC	GG	CC	GG	AA	GG	CC	TT	AC	CT	AC	CT	TT	CC	CT	CC	TT	AA	TT	AG
PA 150	1012Q6	√	GG	AC	GG	CC	GG	AA	GG	CC	TT	AC	CT	AC	CT	TT	CC	CT	CC	TT	AA	TT	AG
PA 150	1010Q6	√	GG	AC	GG	CC	GG	AA	GG	CC	TT	AC	CT	AC	CT	TT	CC	CT	CC	TT	AA	TT	AG
PA 150	1009Q6	Off-type	TT	AC	GG	CC	GG	AA	AA	CT	TT	CC	CT	AA	CC	CT	TT	CC	CT	TT	GG	CT	AA
PA 150	Field D 679 Marper Farm Trinidad	Reference	GG	AC	GG	CC	GG	AA	GG	CC	TT	AC	CT	AC	CT	TT	CC	CT	CC	TT	AA	TT	AG
IMC 76	738Q6	Off-type	TT	AC	TT	CC	AG	AC	AG	CT	CT	CC	CT	AA	CT	CC	TT	CC	CT	GG	AG	CT	AG
IMC 76	737Q6	Off-type	TT	AC	TT	CC	AG	AC	AG	CT	CT	CC	CT	AA	CT	CC	TT	CC	CT	GG	AG	CT	AG
IMC 76	729Q6	√	TT	CC	GT	CC	GG	AC	AG	CT	TT	AC	CC	AC	CC	CC	TT	CC	CC	GT	AG	CT	AG
IMC 76	727Q6	√	TT	CC	GT	CC	GG	AC	AG	CT	TT	AC	CC	AC	CC	CC	TT	CC	CC	GT	AG	CT	AG
IMC 76	724Q6	Off-type	GG	CC	TT	CC	GG	AA	AA	CC	CC	CC	TT	AA	TT	TT	TT	TT	CC	TT	GG	TT	AA
IMC 76	Field D 144 Marper Farm Trinidad	Reference	TT	CC	GT	CC	GG	AC	AG	CT	TT	AC	CC	AC	CC	CC	TT	CC	CC	GT	AG	CT	AG
MAN 15/60	7Q6 e	Off-type	TT	CC	GT	CT	AG	AC	AG	TT	CT	AC	CC	AC	CT	CC	CC	CT	CT	TT	AG	TT	AA
MAN 15/60	85Q6 e	Off-type	TT	AA	GG	CT	AA	AA	AG	TT	TT	CC	CC	AC	CC	CT	TT	CT	TT	TT	AG	CT	GG
MAN 15/60	83Q6 e	√	GT	CC	GT	CC	AG	AC	AA	TT	CT	AC	CC	AC	CT	CC	CC	CT	CT	TT	AG	TT	AA
MAN 15/60	13Q6 e	√	GT	CC	GT	CC	AG	AC	AA	TT	CT	AC	CC	AC	CT	CC	CC	CT	CT	TT	AG	TT	AA
MAN 15/60	46Q6 e	√	GT	CC	GT	CC	AG	AC	AA	TT	CT	AC	CC	AC	CT	CC	CC	CT	CT	TT	AG	TT	AA
MAN 15/60	Field D237 T5 Trinidad	Reference	GT	CC	GT	CC	AG	AC	AA	TT	CT	AC	CC	AC	CT	CC	CC	CT	CT	TT	AG	TT	AA
NA 79	952Q6	√	GT	AC	GG	CC	AA	AA	AG	CT	TT	CC	CT	AA	CT	CC	TT	CC	CC	GT	GG	TT	AA
NA 79	958Q6	Off-type	GG	AC	GG	CC	GG	AA	GG	CC	TT	AC	CT	AC	CT	TT	CC	CT	CC	TT	AA	TT	AG
NA 79	Field D612 Marper farm Trinidad	Reference	GT	AC	GG	CC	AA	AA	AG	CT	TT	CC	CT	AA	CT	CC	TT	CC	CC	GT	GG	TT	AA

Figure 1. Intraclonal mislabeling (off-type) identified in 160 cacao trees from Ghana cacao collections based on 53 SNP markers (of which only 21 loci were presented). The true-to-type clones were marked as "√". The SNP profiles of the reference clones were generated using original trees from International Cacao Genebank, Trinidad.

and of these, two are universal collections (representing nearly all of the known genetic diversity): CATIE (Centro Agronómico Tropical de Investigación y Enseñanza) in Costa Rica and ICG,T in Trinidad and Tobago (Motilal et al, 2013; Wadsworth and Harwood, 2000). Mislabeled plants have been identified as a serious problem in germplasm collections (Hurka et al., 2004). Significant efforts have been made to solve the problem in some international cacao collections (Motilal et al., 2013; Zhang et al., 2009a,b); however, the mislabeling problem in most of the various national collections has not been systematically addressed. Until recently, tools have not been available to clearly identify mislabeled germplasm accessions. Molecular markers such as AFLP (amplified fragment length polymorphism) have sufficient discriminatory power to distinguish cacao accessions; however, these tools often failed to reach clear conclusions, with convincing statistical rigor, that two genotypes are identical (Christopher et al., 1999; Perry et al., 1998; Sounigo et al., 2001).

In the past few years, microsatellite markers have been widely used in cacao genotyping and individual identification, enabling systematic assessment of genetic identity in national and international cacao genebanks (Zhang
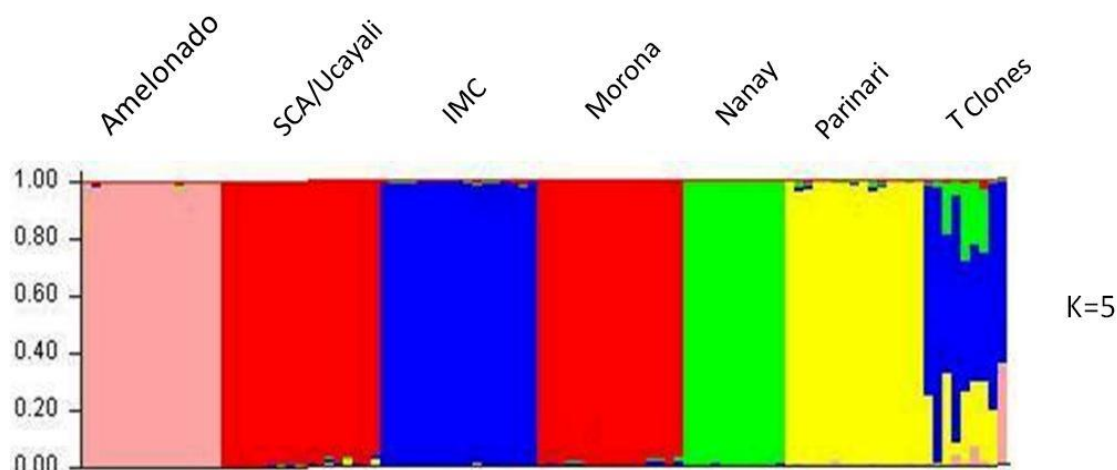
Figure 2. Verification of genetic membership for ten T clones of cacao in Ghana cacao germplasm using assignment test. The computer program STRUCTURE was used, where K is the potential number of genetic clusters that may exist in the overall sample of individuals. Each vertical line represents one individual multilocus genotype. Individuals with multiple colors have admixed genotypes from multiple clusters. Each color represents the most likely ancestry of the cluster from which the genotype or partial genotype was derived. Clusters of individuals are represented by colors.

Table 3. Parentage verification for cacao selections with known breeding pedigree, based on 53 SNP markers with LOD scores at 80 and 95% probability. The SNP profiles of the parental clones were generated using original trees from International Cacao Genebank, Trinidad.

Offspring ID	Recorded Pedigree	Tested candidate mother/father	LOD score*
T16/613	IMC 24 OP	IMC 24 (N/A)	N/A
T16/613	IMC 24 OP	Amelonado 22	8.83
T60/877	PA 7 x NA 32	PA 7	7.97
T60/877	PA 7 x NA 32	NA 32	7.01
T63/967	NA 32 x IMC 67	NA 32	7.13
T63/967	NA 32 x IMC 67	IMC 67 (N/A)	N/A
T63/971	NA 32 x IMC 67	NA 32	6.41
T63/971	NA 32 x IMC 67	IMC 67 (N/A)	N/A
T65/238	PA 7 x IMC 47	PA 7	2.34
T65/238	PA 7 x IMC 47	IMC 47	3.89
T65/326	PA 7 x IMC 47	PA 7	5.17
T65/326	PA 7 x IMC 47	IMC 47	6.08
T79/501(a)	NA 32 x PA 7	PA 7	0.04
T79/501(a)	NA 32 x PA 7	NA 32	0.45
T79/501(b)	NA 32 x PA 7	PA 7	6.65
T79/501(b)	NA 32 x PA 7	NA 32	7.32
T85/799	IMC 60 x NA 34	IMC 60 (N/A)	N/A
T85/799	IMC 60 x NA 34	NA 34	3.04

*Critical LOD (the natural logarithm of the likelihood) ratio for assignment of maternity and paternity are 5.70 at >95% confidence and 2.75 at >80% confidence.

et al., 2009a; Motilal et al., 2009,2010). In contrast to dominant markers, identical genotypes can have a 100% match in the multilocus SSR profiles without ambiguity, thus accuracy of identification is significantly improved. Reference SSR profiles of cacao clones have been deposited in the International Cacao Germplasm Database at the University of Reading, UK (http://www.icgd.rdg.ac.uk/index.php). However, comparison of genotyping results from different laboratories has not been straight forward. The effectiveness of clone identification via SSR fingerprints depends on the number of loci used for genotyping, as well as the rate of geno-

typing error. For example, it may require multiple repeated genotyping runs to reach the "consensus genotype". Moreover, data generated from different genotyping platforms can be difficult to compare with one another because the same allele may be binned differently, leading to false conclusions.

The present study demonstrated that using the SNP-based multilocus fingerprints significantly improved the efficiency of genotype identification. Off-type identification, through the comparison with reference SNP profiles, is straightforward when reference trees are available. The reference trees used in the present study were sampled from the original collections maintained at Marper Farm and San Juan Estate in Trinidad, and Cabiria Farm, CATIE, in Costa Rica. These reference trees have been genotyped by SSR markers and passed through rigorous statistical population genetics tests (Motamayor et al., 2008; Zhang et al., 2009a,b; Johnson et al., 2009).

Parentage verification and assignment test

Many national cacao germplasm collections also maintain local varieties and breeding lines, which do not have a reference tree in international germplasm collections. In this situation, indirect verification such as Bayesian assignment test, parentage analysis, and sibship reconstruction need to be applied. The present study demonstrated how parentage analysis and Bayesian assignment test can be used to verify the genetic identity and pedigree information. Of the eight tested accessions, six were confirmed to have the correct maternal or paternal parent matching with the breeding record. Among them, T63/967 and T63/971 were supposed to be siblings and their verified parentage supported each other. T16/63 was recorded as the open pollinated progeny of IMC 24. Parentage analysis identified Amelonado 22, at a 95% confidence level, as the hidden pollen parent. For candidate parents that did not reach the 80% confidence level, the failure indicates mislabeling (off-type). Another possibility is possible conta-mination due to unwanted pollen or self-compatibility.

The SCA/Ucayali and Morona accessions represent two distinct geographical regions and were clustered as two different genetic groups when SSR markers were used (Zhang et al., 2009b; Motamayor et al., 2008). However, in the present study, the Bayesian clustering analysis based on 53 SNP markers did not significantly differentiate these two germplasm groups (Figure 2). Differences in genetic distances quantified by SNP and SSR markers have been reported in other crops. Yang et al. (2011) reported a correlation between kinship coefficient estimated by SSR and SNP of 0.69 in maize. Murray et al. (2009) found that some sorghum individuals shifted groups, depending upon whether SSR or SNP data was used in the STRUCTURE program. The discrepancy in stratification based on the two marker systems could also be due to the relatively small number of SNP markers

used in the present study. Yu et al. (2009) showed that kinship estimated using 1,000 SNPs was consistent with that estimated with 100 SSRs in maize. Van Inghelandt et al. (2010) proposed that 7 to 11 times more SNPs than SSR markers should be used for analyzing population structure and genetic diversity in maize germplasm. Given that our previous stratification was based on 15 SSR markers, it would require more than 100 SNP markers to reach the same precision level. Additional SNP markers need to be evaluated for cacao and the correlation between SNP markers and SSR markers needs to be systematically assessed.

In addition to the limitation due to a limited number of SNP markers, the discrepancy between the two marker systems might also be partially explained by the derivation of the SNP markers used in the present study from the EST data. A set of unequivocally neutral SNP markers would be ideal. Despite the lack of differentiation between the SCA/Ucayali and Morona populations, the assignment test correctly excluded both groups in terms of parentage contribution to the tested T clones. The assignment of the T clones is fully consistent with the outcome of parentage analysis and is consistent with the recorded pedigree (Lockwood and Gyamfi, 1979). The high repeatability of the genotyping result, as demonstrated by the multiple trees for some cacao germplasm maintained in the Ghana collection, as well as the consistency in pedigree records and parentage analysis, demonstrated that these SNP markers provide a reliable and efficient solution for cacao genotype identification. This modest set of SNP markers thus constitutes a cost-effective marker resource, suitable for backstopping large-scale clone propagation in cacao. Nonetheless, the study also showed that a larger number of SNP markers would be needed for comprehensive diversity analysis.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

Adu-Ampomah Y, Adomako B, Opoku IY (2006). Cocoa population breeding approaches in Ghana. In: Eskes AB, Efron Y (eds) Global Approaches to Cocoa Germplasm Utilization and Conservation. Final

report of the CFC/ICCO/IPGRI project on Cocoa Germplasm Utilization and Conservation: a Global Approach (1998-2004), CFC, Amsterdam, The Netherlands/ICCO, London, UK/IPGRI, Rome, Italy, pp 41-46.

Aikpokpodion PO, Adetimirin VO, Ingelbrecht I, Schnell RJ, Kolesnikova-Allen M (2005). Assessment of genetic diversity of cacao, *Theobroma cacao* L., collections in Nigeria using simple sequence repeat markers. In: Denamany G, Lamin K, Ling A, Maisin N, Ahmad, AC, Saripah B, Nuraziawati MY (eds) Sustainable cocoa economy through increase in productivity, efficiency and quality: Proceedings of 4th Malaysian International Cocoa Conference, Kuala Lumpur, Malaysia, 18th-19th July 2005, Malaysian Cocoa Board, Kota Kinubalu, pp 83-86.

Allegre M, Argout X, Boccara M, Fouet O, Roguet Y, Bérard A, Thévenin JM, Chauveau A, Rivallan R, Clement D, Courtois B, Gramacho K, Boland-Augé A, Tahi M, Umaharan P, Brunel D, Lanaud C (2012). Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L. DNA Res. 19:23-35.

Argout X, Fouet O, Wincker P et al. (2008). Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* generated from various tissues and under various conditions. BMC Genomics 9:512.

Bartley BGD (2005). The genetic diversity of cacao and its utilization. CAB International, CABI Publishing, Wallingford, Oxfordshire.

Buckler ES, Thornsberry J (2002). Plant molecular diversity and applications to genomics. Curr. Opin. Plant Biol. 5:107-111.

Christopher Y, Mooleedhar V, Bekele F, Hosein F (1999). Verification of accession in the ICG, T using botanical descriptors and RAPD analysis. In: Annual Report 1998, Cocoa Research Unit, The University of the West Indies, St. Augustine, Trinidad and Tobago. pp. 15-18.

Cuatrecasas J (1964). Cacao and its allies: A taxonomic revision of the genus *Theobroma*. Contributions from the United States National Herbarium Volume 35, part 6, Washington DC, Smithsonian Institution Press, Washington DC. pp. 375-614.

Doyle JJ, Doyle JL (1990), Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Edwin J, Masters WA (2005). Genetic improvement and cocoa yields in Ghana. Exp. Agric. 41:491-503.

Efombagan IB, Motamayor JC, Sounigo O, Eskes AB, Nyasse S, Cilas C, Schnell RJ, Manzanares-Dauleux M, Kolesnikova-Allen M (2008). Genetic diversity and structure of farm and genebank accessions of cacao (*Theobroma cacao* L.) in Cameroon revealed by microsatellite markers. Tree Genet. Genomes 4:821-831.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611-2620.

Evett IW, Weir BS (1998). Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists. Sinauer, Sunderland, Massachusetts, USA.

FAOSTAT. Food and Agricultural commodities production. FAO statistical databases. Available from: http://faostat3.fao.org/home/index.html.

Figueira A, Janick J, Levy M, Goldsbrough P (1994). Reexamining the classification of *Theobroma cacao* L. using molecular markers. J. Am. Soc. Hortic. Sci. 119:1073-1082.

Guiltinan M, Verica J, Zhang D, Figueira A (2008). Genomics of *Theobroma cacao*, the Food of the Gods. In: Moore P, Ming R (eds) Genomics of Tropical Crop Plants, Springer, New York. pp. 145-170.

Hurka H, Neuffer B, Friesen N (2004). Plant genetic resources in botanical gardens. In: Forkmann G, Michaelis S (eds) Proceedings of the 21st International Symposium on Breeding Ornamentals, Part II. Acta Hortic. 651:35-44.

Ji K, Zhang DP, Motilal L, Boccara M, Lachenaud P, Meinhardt LW (2013). Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. Genet. Resour. Crop Evol. 60:441-453.

Johnson ES, Bekele FL, Brown SJ, Song Q, Zhang D, Meinhardt LW, Schnell RJ (2009). Population structure and genetic diversity of the

Trinitario cacao (*Theobroma cacao* L.) from Trinidad and Tobago. Crop Sci. 49:564-572.

Kalinowski ST, Taper ML, Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16:1099-1006.

Kuhn DN, Livingstone D, Main D, Zheng P, Saski C, Feltus FA, Mockaitis K, Farmer AD, May GD, Schnell RJ, Motamayor JC (2012). Identification and mapping of conserved ortholog set (COS) II sequences of cacao and their conversion to SNP markers for marker-assisted selection in *Theobroma cacao* and comparative genomics studies. Tree Genet. Genomes 8:97-111.

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda PJL (1999). Isolation and characterization of microsatellites in *Theobroma cacao* L. Mol. Ecol. 8:2141-2143.

Lerceteau E, Robert T, Pétiard V, Crouzillat D (1997). Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD and RFLP markers. Theor. Appl. Genet. 95:10-19.

Livingstone DS, Freeman B, Motamayor JC, Schnell RJ, Royaert S, Takrama J, Meerow AW, Kuhn DN (2012). Optimization of a SNP assay for genotyping *Theobroma cacao* under field conditions. Mol. Breed. 30:33-52.

Lockwood G, Gyamfi MMO (1979). The CRIG cocoa germplasm collection with notes on codes used in the breeding programme at Tafo and elsewhere. Tech. Bull. 10, Cocoa Research Institute, Ghana, 62 pp.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7:639-655.

Motamayor JC, Lachenaud P, da Silva e Mota JW, Loor G, Kuhn DN, Brown JS, Schnell RJ (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree. PLoS ONE 3:e3311. 10.1371/journal.pone.0003311.

Motilal L, Zhang D, Umaharan P, Mischke S, Mooleedhar V, Meinhardt LW (2010). The relic Criollo cacao in Belize- genetic diversity and relationship with Trinitario and other cacao clones held in the International Cocoa Genebank, Trinidad. Plant. Genet. Resour. 8:106-110.

Motilal LA (2004). The potential of cacao microsatellites amplification across diverse plant taxa. In: Thangadurai D, Pullaiah T, Balatti PA (eds) Genetic Resources and Biotechnology, Vol. 2, , Regency Publications, New Delhi, India. pp. 24-49.

Motilal LA, Butler D (2003). Verification of identities in global cacao germplasm collections. Genet. Resour. Crop. Ev. 50:799-807.

Motilal LA, Zhang D, Mischke S, Meinhardt LW, Umaharan P (2013). Microsatellite-aided detection of genetic redundancy improves management of the International Cocoa Genebank, Trinidad. Tree. Genet. Genomes 9:1395-1411.

Motilal LA, Zhang D, Umaharan P, Mischke S, Boccara M, Pinney S (2009). Increasing accuracy and throughput in large-scale microsatellite fingerprinting of cacao field germplasm collections. Trop. Plant. Biol. 2:23-27.

Murray SC, Rooney WL, Hamblin MT, Mitchell SE, Kresovich S (2009). Sweet sorghum genetic diversity and association mapping for brix and height. Plant Genome 2:48-62.

N'Goran JAK, Laurent V, Risterucci AM, Lanaud C (2000). The genetic structure of cacao populations (*Theobroma cacao* L.) revealed by RFLP analysis. Euphytica 115:83-90.

Peakall R, Smouse PE (2006). Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288-295.

Peakall R, Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537-2539.

Perry MD, Davey MR, Power JB, Lowe KC, Bligh HFJ, Roach PS, Jones C (1998). DNA isolation and AFLP genetic fingerprinting of *Theobroma cacao* L. Plant Mol. Biol. Rep. 16:49-59.

Posnette AF (1986). Fifty years of cocoa research in Trinidad and Tobago. Cocoa Research Unit, University of the West Indies, St. Augustine, Trinidad, 131 pp.

Powis TG, Cyphers A, Gaikwad NW, Grivetti L, Cheong K (2011). Cacao use and the San Lorenzo Olmec. Proc. Natl. Acad. Sci. USA 108:8595-8600.

Pritchard JK, Stephens M, Donelly P (2000). Inference of population structure from multilocus genotype data. Genetics 155:945-959.

Sounigo O, Christopher Y, Bekele F, Mooleedhar V, Hosein F (2001). The detection of mislabelled trees in the International Cocoa Genebank, Trinidad (ICG,T). In: Proceedings of the International Workshop on the New Technologies and Cocoa Breeding, 16th–17th October 2000, Kota Kinabalu, Sabah, Malaysia, INGENIC (International Group for Genetic Improvement of Cocoa). pp. 34-39.

Sounigo O, Umaharan R, Christopher Y, Sankar A, Ramdahin S (2005). Assessing the genetic diversity in the International Cocoa Genebank, Trinidad (ICG,T) using isozyme electrophoresis and RAPD. Genet. Resour. Crop. Evol. 52:1111-1120.

Sounigo, O, Risterucci A-M, Clement D, Fouet O, Lanaud C (2006). Identification of off-types of clones used in the International Clone Trial using DNA analyses. In: Eskes AB, Efron Y (eds) Global approaches to cocoa germplasm utilization and conservation. Final report of the CFC/ICCO/IPGRI project on Cocoa Germplasm Utilization and Conservation: a Global Approach (1998-2004), CFC, Amsterdam, The Netherlands/ICCO, London, UK/IPGRI, Rome, Italy, pp. 82-86.

Takrama J, Dadzie AM, Opoku FK, Padi FK, Adomako B, Asu-Ampomah Y, Livingstone DS, Motamayor JC, Schnell RJ, Kuhn RJ (2012). Applying SNP marker technology in the cacao breeding programme in Ghana. Afr. Crop. Sci. J. 20:67-75.

Takrama JF, Cervantes-Martinez C, Phillips-Mora W, Brown JS, Motamayor JC, Schnell RJ (2005). Determination of off-types in a cocoa breeding programme using microsatellites. INGENIC Newsletter 10:2-8.

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor. Appl. Genet. 120:1289-1299.

Wadsworth RM, Harwood T (2000). International Cocoa Germplasm Database, ICGD 2000 V4.1. London International Financial Futures and Options Exchange and the University of Reading, UK.

Waits LP, Luikart G, Taberlet P (2001). Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol. Ecol. 10:249-256.

Whitkus R, de la Cruz M, Mota-Bravo L, Gómez-Pompa A (1998). Genetic diversity and relationships of cacao (Theobroma cacao L.) in southern Mexico. Theor. Appl. Genet. 96:621-627.

Wood GAR, Lass RA (1985). Cocoa. 4th Edn, Essex: Longman Group Ltd, 620 pp.

Yang XH, Xu YB, Shah T, Li HH, Han ZH, Li JS, Yan JB (2011). Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. Genetica 139:1045-1054.

Yu JM, Zhang ZW, Zhu CS, Tabanao DA, Pressoir G, Tuinstra MR, Kresovich S, Todhunter RJ, Buckler ES (2009). Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping. Plant Genome 2:63-77.

Zhang D, Boccara M, Motilal L, Mischke S, Johnson ES, Butler D, Bailey BA, Meinhardt, LW (2009b). Molecular characterization of an earliest cacao (Theobroma cacao L.) collection from Peruvian Amazon using microsatllite DNA markers. Tree Genet. Genomes 5:595-607.

Zhang D, Mischke S, Johnson ES, Mora A, Phillips-Mora W, Meinhardt LW (2009a). Molecular characterization of an International cacao collection using microsatellite markers. Tree Genet. Genomes 5:1-10.

Full Length Research Paper

Characterization and *in vitro* studies on anticancer activity of exopolymer of *Bacillus thuringiensis* S13

K. Parthiban[1,2], V. Vignesh[2] and R. Thirumurugan[2*]

[1]Department of Microbiology, Hajee Karutha Rowther Howdia College (Autonomous) Uthamapalayam - 625533, Tamilnadu, India.
[2]Department of Animal science, School of Life Sciences, Bharathidasan University, Tiruchirappalli - 620024, Tamilnadu, India.

The objective of the present work was to isolate and characterize the exopolymer producer from a marine environment. The exopolymer producing strain (S13) was identified as *Bacillus thuringiensis* S13. Characterization of exopolymer shows the presence of brominated compound responsible for cytotoxicity on lung cancer cell line (A549) on XTT assay. An *in vitro* study of bacterial exopolymer shows the presence of cytotoxic effects on cell lines. Further, active compound in exopolymer responsible for cytotoxicity has to be characterized. The exopolymer produced by *B. thuringiensis* S13, showed potent cytotoxic effects, and could be used as therapy in cancer after further studies.

Key words: *Bacillus thuringiensis* S13, exopolymer, IR, GCMS analysis and lung cancer cell line.

INTRODUCTION

Ocean inhabits millions of unexplored organisms, which secretes valuable unexplored compounds. Up to 2008, only about ten thousand compounds have been evolved from marine organisms. The assorted environments in marine, force the bacteria to secrete some compounds to survive (Jensen and Fenical, 1994). In the marine ecosystem, bacteria secrete many bioactive exopolymers compared with another local, since marine has numerous and distinct capabilities to endure (Annarita et al., 2010). Majority of bacterial species produced secondary metabolites as exopolymer containing sugar and non-sugar components like amino sugars, protein

(Sutherland, 1977) and fatty acids (Hayashida-Soiza et al., 2008). The adaptation of bacteria to diverse habitat can determine the production of unique secondary metabolites as exocellular. Sea provides an abundant source of nutrients and other parameters; the bacteria on the sea surface aggregates to form biofilm (Ng and Hu, 1989; Donlan and Costerton, 2002). Some bacteria do not produce any compounds because they make contact with other bacterial species or extracellular products from other bacteria. Several investigation exhibits number of biologically active compounds from marine bacteria (Barsby et al., 2001) showing antimicrobial activities,

*Corresponding author. E–mail: ramthiru72@gmail.com. Tel: +91 431 2407040. Fax: +91 431 2407045.

antiviral, immunosuppressive, enzyme inhibitor metabolites, receptor antagonistic, antitumor activities and anticoagulant properties (Reichenbach, 2001). The chemical nature of bacterial exopolymer is often highly complex. It was found that most of the marine isolates exhibit antagonistic activity. A term cancer is used to define uncontrolled cell growth. The cells divide rapidly, forming malignant tumour, and invading the adjacent cells. Cancer affects various parts of the body. Each cancer is different, and the chance of surviving depends on the cancer and site of cancerous growth. Cancer can affect any human being regardless of age. About 13% of human death is due to cancer throughout the world. Lung cancer causes more people to die than other cancers worldwide, accounting 1.3 million people to die annually (Khuri et al., 2001). At present, the clinical responses to patients of effectively targeted therapies for lung cancer are still inadequate (Danesi et al., 2003). Hence, the development of new effective anti-cancer drugs for lung cancer is the most important need of the day.

Polysaccharides derived from a microorganism have specific broad ranged properties such as antitumor, antioxidant activity. Bacterial exopolymeric substances have a wide range of applications (Adriana et al., 2005; Chen et al., 2008) that makes them interesting from the biotechnological point of view.

MATERIALS AND METHODS

Screening and identification

Soil samples were collected from Mandapam, coastal area (790 8' E, 9° 17' N) of south India at a depth of 5 m. The bacterial strains were screened in Zobell agar plates (Himedia India) followed by serial dilution method. The plates were incubated at 25°C for a period of seven days. Morphology of bacterial colonies were observed and selected based on their mucoid morphology (Ng and Hu, 1989). After isolation, the strain (S13) was characterized phenotypically and biochemically, which was subcultured and maintained in 20% glycerol at 4°C as stock (Jean-Marc et al., 1990). For 16S rDNA gene sequencing, the bacterial colonies were picked and suspended in 0.5 µl of sterilized saline and centrifuged at 10,000 rpm for 10 min. After removal of the supernatant, the pellet was suspended in 0.5 µl of InstaGene Matrix (Bio-Rad, USA) and incubated at 56°C for 30 min, then heated at 100°C for 10 min. After heating, supernatant can be used for PCR. Template DNA (1 µl) was treated with 20 µl of PCR reaction solution, and 27F/1492R primers were used for amplification. Unincorporated PCR primers and dNTPs from PCR products were removed by using Montage PCR clean up kit (Millipore). The 16S rDNA was amplified by polymerase chain reaction (PCR) using the primers 518F CCAgCAgCCgCggTAATACg and 800R TACCAgggTATCTAATCC universal primers. The purified PCR products were sequenced by using 2 primers as described earlier. Sequencing was performed using Big Dye terminator cycle sequencing kit (Applied BioSystems, USA). Sequencing products were resolved on automated DNA sequencing system. DNA sequence was obtained using DNA sequencer and the PCR products were sequenced using the same PCR primers and other internal primers to confirm the sequence (Dereeper et al., 2008).

The sequence of 16s rDNA gene was compared with the sequence data through BLAST searching tool in NCBI. As a result of BLAST search, the sequences were aligned to their closest sequence with MUSCLE (v3.7) and refined using Gblocks 0.91b (Castresana, 2000). The phylogenetic tree was constructed by SH-like method implemented in the PhyML program.

Extraction and characterization exopolymeric substance

The isolate S13 was grown in Zobell marine broth for 7 days at 32°C. The exopolymer was extracted, treated with two volumes of 95% cold ethanol to the cell free supernatant and stored at 4°C for 24 h. The precipitate was collected by centrifugation and washed with distilled water. The precipitate was redissolved in distilled water and dialyzed (molecular weight (mol. wt.) cut off 8000 dalton) against distilled water for 2 days to remove the salts from exopolymer. The weight of the exopolymer was measured after drying at 45°C (Bragadeeswaran et al., 2011).

Characterization of bacterial exopolymeric substance

The total sugar content was measured using phenol sulphuric acid assay with glucose as standard (Bruckner, 1955). The total protein content was measured by Lowry's method with bovine serum albumin as standard (Lowry et al., 1951). FT-IR spectra were recorded for exopolymer with a resolution of 4 cm^{-1} in the 4000-400 cm^{-1} region (Lijour et al., 1994). GCMS analysis was performed in JEOL GCMATE II GCMS (Mancuso et al., 2004). Helium was the carrier gas. Peaks were identified by comparing with known standards.

Optimization

The optimum carbon source for the growth and exopolymer production was achieved by fermentation in 100 ml of basal salt media. The different concentration (0.5, 1, 1.5, 2 and 2.5%) of carbon sources (glucose, sucrose, lactose and galactose) (Mata et al., 2008) were tested by inoculation with 2 ml of 24 h culture. The basal salt medium was supplied with 1% of different nitrogen source (peptone, yeast extract, ammonium chloride, and ammonium nitrate) to determining the effect of nitrogen source (Sung-Hwan et al., 2000). The isolate, which exhibited maximum growth and yield of exopolymer for carbon and nitrogen sources respectively, was selected for optimization of carbon: Nitrogen ratio. The basal salt medium with carbon source and nitrogen source which showed maximum yield were selected and different concentration of nitrogen was supplied (0.05 to 0.1%) for the respective isolate (Read and Costerton, 1987). The optimum pH for exopolymer synthesis was determined in basal salt media with different pH (5, 5.5, 6, 6.5, and 7) and, exopolymer production was achieved in production media (casein hydrolysate media) at 37°C in a shaker. The growth rate was measured at 540 nm.

In vitro anticancer activity

Briefly, the cancer cell lines (A549 lung cancer) were procured from National Centre for Cell Science, Pune. The cell lines were grown and maintained in Minimal essential medium (MEM, GIBCO) supplemented with 4.5 g/L glucose, 2 mM L-glutamine and 5% fetal bovine serum at 37°C in 5% CO_2 incubator. From the T-25 flask, the trypsinized cells were seeded in each well of 96 well flat-bottomed plates and incubated in 5% CO_2 at 37°C. After 24 h of incubation, the supernatant was discarded and, the cells were
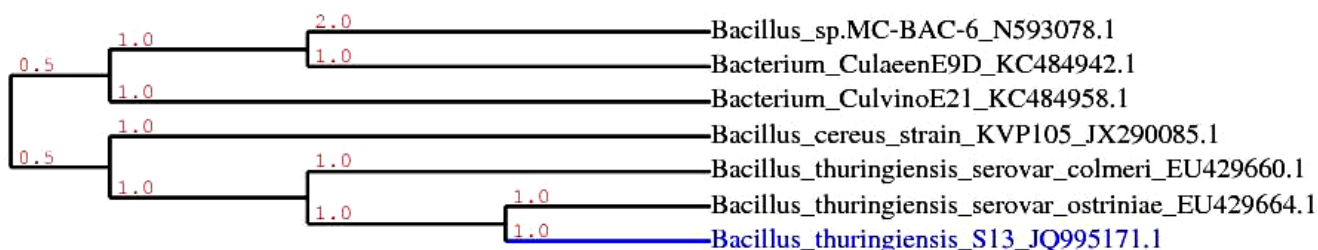
Figure 1. Phylogenetic tree of isolate S13 to their closest NCBI (BLASTn) strains based on the 16S rDNA gene sequences.

pretreated with growth medium. Subsequently, it was mixed with different concentrations (12.5, 25, 50, 100 and 200 µg/ml) of exopolymer and cisplatin (an anticancer drug as standard) separately in triplicates to achieve a final volume of 100 µl and then cultured for 48 h. The exopolymer and cisplatin were prepared separately as 1.0 mg/ml concentration stock solutions in dimethyl sulfoxide (DMSO). Culture medium and solvent were used as blank. Each well then received 50 µl of fresh XTT (0.9 mg/ml in RPMI along with XTT activator reagent) followed by incubation for 2 h at 37°C. At the end of incubation, 96 micro well plates were shaken for 15 s (Roehm et al., 1991; Stevens and Olsen, 1993).

RESULTS

The exopolymer producing strain was isolated from marine sediment; it forms circular convex mucoid colonies in Zobell agar medium. The bacteria was characterized as Gram positive; rod shaped showing positive results to VP, catalase, oxidase, protein and starch hydrolysis reactions, negative for indole, MR, citrate utilization test. The isolate produce endospore and showed the presence of crystals. The 16S rRNA sequence revealed the isolate was *Bacillus thuringiensis.* The sequence was submitted in Genbank NCBI, and the accession number assigned as JQ995171.1. Figure 1 shows the phylogenetic relationship of *B. thuringiensis* S13 with its closely related sequence of blast results. From the results, it was confirmed that, it belonged to the Phylum *Firmicutes* and family *Bacillaceae.* Calorimetric study showed 62.01% of sugar and 3.27% of protein in exopolymer. FT IR spectrum (Figure 2.) revealed characteristic functional groups of exopolymer; C-Br stretching noticed at 605.61 and 651.89 cm^{-1}. Four or more CH_2 groups occur in an open chain at 752.19 cm^{-1}. Further, aryl fluoride C-F stretching peak was noticed at 1195.78 cm^{-1}. A broad stretching of CH_3, CH_2-CH_3 at 1334.65 to1454.23 cm^{-1}. Stretching peak at 1668 and 1334 cm^{-1} indicates the presence of COOH group. C-H stretching was noticed in 2885.31 to 2974.03 cm^{-1}, O-H stretching was observed in the range of 3195.83 to 3313.48 cm^{-1}.

Gas chromatography–mass spectrometry (GC-MS) analysis showed the presence of saturated fatty acid and brominated compounds (Table 1). All sugars support the growth and exopolymer production. As far as the carbon source is concerned, the growth and production rate of *B. thuringiensis* S13 is directly proportional to the concentra-

tion of sugars. For all sugars, the growth and production rate dramatically increased for sugar concentration. However, the highest yield (58.5 mg/l) of exopolymer was observed when glucose was supplied as a carbon source (Figure 3). Of the various nitrogen source tested, peptone produced the highest yield of exopolymer (peptone 49.5 mg/l, yeast extract 47.2 mg/l, ammonium chloride 41.5 mg/l, and ammonium nitrate 40 mg/l). The optimum carbon and nitrogen concentration for exopolymer production was noticed when 0.6% peptone was supplemented with 2.5% of glucose (121 mg/l) (Table 2). At different pH, the cultures showed a sigmoid growth (Figure 4). The effect of pH (6.0, 6.5, 7.0, 7.5 and 8.0) on growth and production was assessed and summarized in Table 3. The isolate showed sigmoidal growth and production at variable pH. However, the significant amount of exopolymer production was observed at pH 7.0 (292 mg/l). At low or high pH, the exopolymer production was stunned.

XTT assay showed that the bacterial exopolymer has an anti-proliferative activity on A549 lung cancer cells. After the addition of different concentration of exopolymer and cisplatin (control) to the cancer cells separately, the growth rate was measured by its optical density value at 490 nm followed by the addition of XTT. It was interesting to note that the exopolymer inhibits 50% of cancer cell growth at a concentration (100 µg/ml) (Table 4). From the results, the half maximal inhibitory concentration (IC50) of the bacterial exopolymer was estimated at 133.27 µg/mL, and IC 50 value for cisplatin at 23.7 µg/mL. Metabolically active cancer cells potentially reduced and convert the yellow tetrazolium salt (XTT) into orange formasan dye (Figure 5a) by the enzyme mitochondrial oxidoreductases. This conversion occurs only in the live cells. However, the bacterial exopolymer treated cells fail to reduce (Figure 5b) indicating that the bacterial exopolymer can inhibits the cell proliferation.

DISCUSSION

The promising strain was selected based on the development of mucoid morphology because it was one of the fundamental screenings for isolation of exopolymer producing bacteria. The presence of mucoid morphology is one of the key factors to screen the exopolymer-produ-
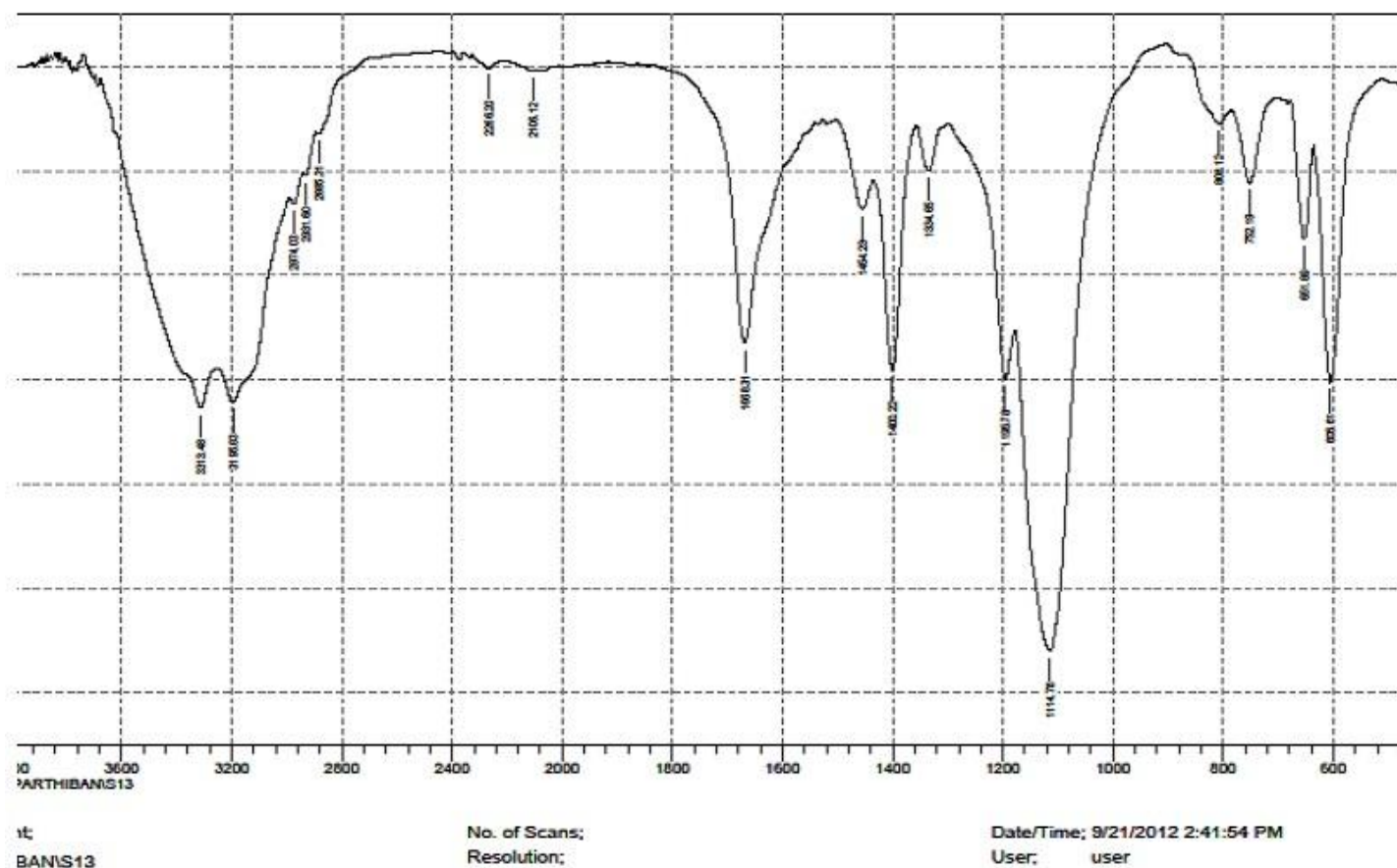
Figure 2. FTIR spectrum of exopolymer. **Note:** Stretching at 752.19 cm^{-1} indicates the presence of CH$_2$ groups and 1668 and 1334 cm^{-1} indicates the presence of COOH group. C-H stretching at 2885.31 to 2974.03 cm^{-1} and O-H stretching was observed in the range of 3195.83 to 3313.48 cm^{-1}.

Table 1. GCMS analysis of non carbohydrate fraction of bacterial exopolymer.

Rt	Name of compound	Mass	Molecular formula
12.58	Pentanoicacid,5-hrdroxy,2,4-di-t-butylphenyl esters	306.4397	C$_{19}$H$_{30}$O$_3$
17.65	1,1,3,1"-Terphenyl,3,3,5,5-tetrabromo-5-{3,5-dibromophenyl}	780	C24H12Br6
23.27	Docosanoic acid,1,2,3-propanetriyl ester	1059.81	C69-H134-O6

cing bacteria (Inmaculada et al., 2010). The exopolymer producing strain was identified by standard morphological, physiological and biochemical methods, compared with Bergey's manual of determinative bacteriology (Holt et al., 1994). Al-Nahas et al. (2011) also reported that the exopolymer production increased, when glucose was supplied as a carbon source in *Pseudoalteromonas* sp. Sutherland (1977) and Cerning et al. (1994) also found that excess of sugar in the culture medium stimulates the production of exopolymer by lactic acid bacteria. In

accordance with this, it is a recognized fact that a similar association exists in the concentration of sugars and the exopolymer production. Kim et al. (2008) in *Weissella hellenica* reported a similar result. The exopolymer synthesis started during the late exponential growth phase and reached the maximum after 3 to 5 days. According to Gancel et al. (1994), the subsequent decrease in the production was probably due to various effects such as physiological stress, enzymatic degradation. Cerning et al. (1994) and West et al. (1998) found similar observation
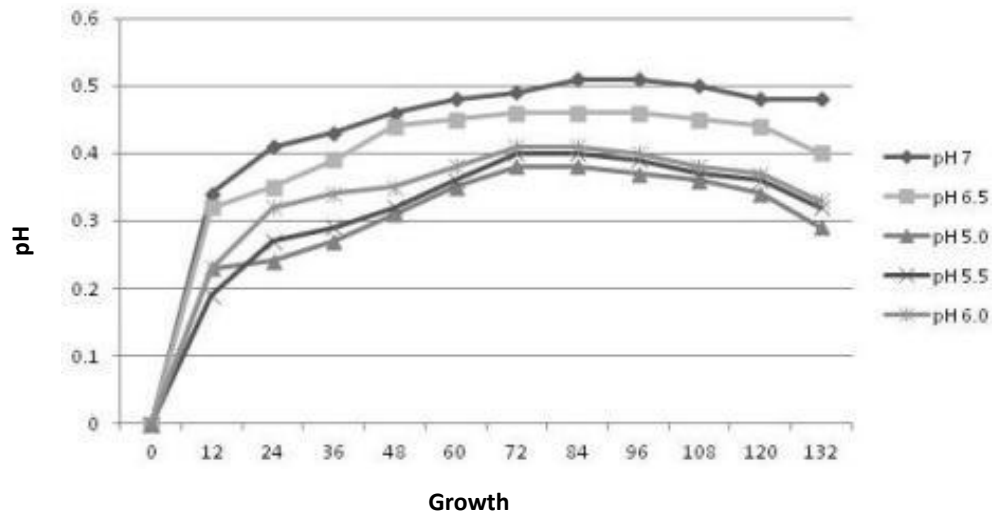
Figure 3. Effect of pH on growth of *Bacillus thuringiensis* S13. **Note:** The isolate exhibits sigmoidal growth and exopolymer production at pH 7 on compared with other pH condition.

Table 2. Effect of carbon nitrogen concentration on exopolymer production.

C:N Concentration		Exopolymer (mg/l)
	Peptone (%)	
	0.5	98.33 ±0.28
Glucose 2.5%	0.6	121.1 ±0.28
	0.7	84.33 ±0.28
+	0.8	65.33 ±0.11
	0.9	59.16 ±0.05
	1	41.33 ±0.28

The isolate produce highest amount of exopolymer on supplying 0.5% of peptone with 2.5% glucose. *Values are the means ± standard deviations of triplicate measurements. Effect of C:N on production of exopolymer.
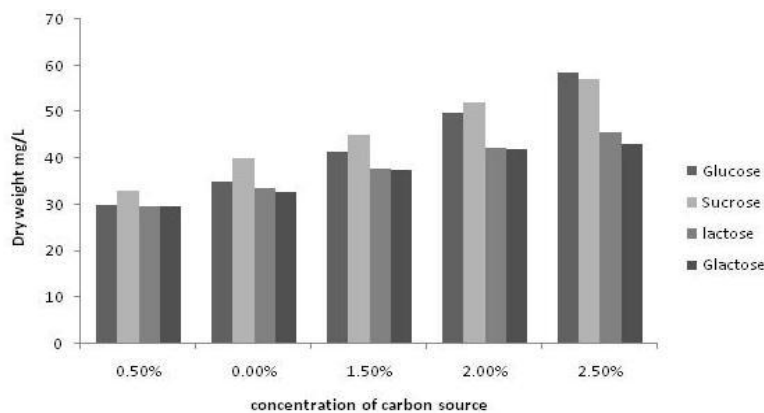
Figure 4. Effect of carbon source on production of exopolymer. **Note:** The isolate S13 produce highest amount of exopolymer when glucose was supplied as carbon source. The exopolymer production increased with the concentration of sugars.

**Table 3.** Effect of pH on exopolymer production.

Bacillus thuringiensis S13	
pH of the media	Dry weight of EPS (mg/l)
pH 5	040
pH 5.5	096
pH 6	134
pH 6.5	198
pH 7	292

The optimum pH for the highest yield of exopolymer is 7.

Table 4. Dose response of bacterial exopolymer on A549 (Lung cancer).

Concentration (ug/ml)	OD of cisplatin (STD) treated cells at 490 nm	% Cell survival	% Cell inhibition	OD of exopolymer treated cells at 490 nm	% Cell survival	% Cell inhibition
12.5	1.04±0.00	75.46±0.05	24.53±0.05	0.68±0.00	99.76±0.05	0.23±0.05
25	0.68±0.00	48.56±0.05	51.43±0.05	0.57±0.00	82.63±0.05	17.36±0.05
50	0.42±0.00	28.63±0.05	71.36±0.05	0.43±0.00	61.23±0.11	38.76±0.11
100	0.26±0.00	16.56±0.05	83.43±0.05	0.36±0.00	50.33±0.57	49.66±0.57
200	0.15±0.00	8.86±0.05	91.13±0.05	0.28±0.00	37.56±0.05	62.43±0.05

The half-inhibitory concentration of bacterial exopolymer against cancer cell is 50 (ug/ml), where as for control 25 (ug/ml). *Values are the means ± standard deviations of triplicate measurements. % = percentage.
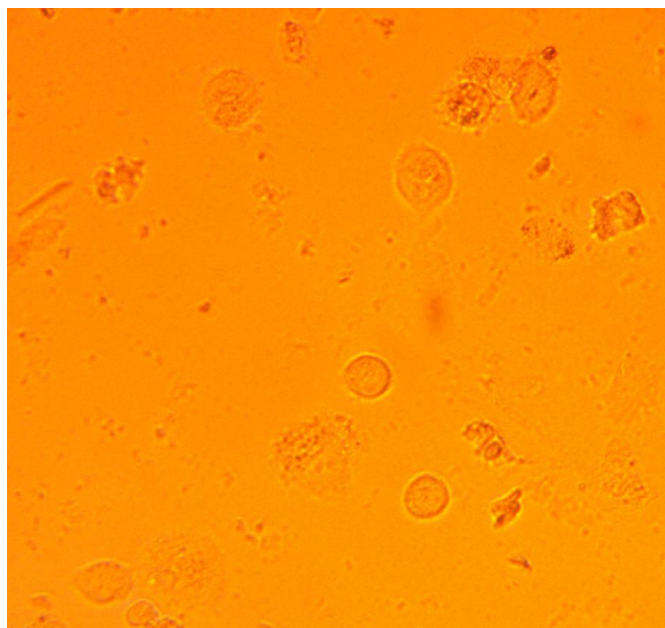
Figure 5a. A549 cell line treated with exopolymer.
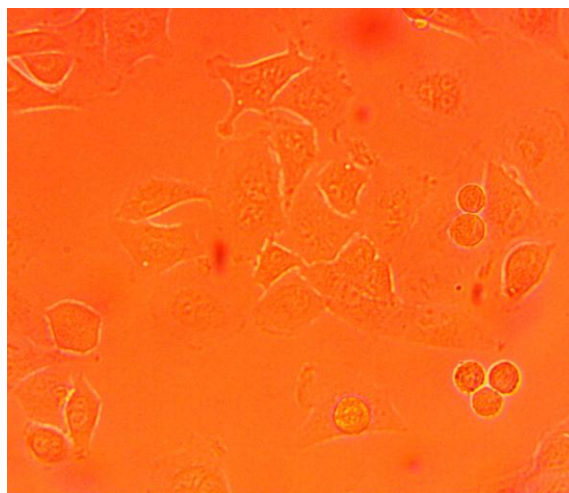
Figure 5b. A549 cell line. **Note:** The color change from yellow to orange (Figure 5b) indicates the reduction of yellow tetrazolium salt (XTT) into orange formazan dye by the metabolically active mitochondrial enzymes, meanwhile, the exopolymer treated cell lines fails to convert (Figure 5a) indicates that the cell become lyses.

in *Lacobacillus casei* and *Sphingomonas pancimobilis* that yielded high quantity of exopolymer when glucose is use as a carbon source. The isolate produced a maximum

amount of exopolymer at pH 7. This result is consistent with Vincent et al. (1994) who studied in *Alvinella pompejana,* which produces the highest amount of exopoly-

mer at pH 7.0 to 7.3. Fourier transform infrared (FTIR) spectrum was relevant to the peaks already has been reported by Lungmann et al. (2007); Braissant et al. (2007); Vijayabaskar et al. (2011); Sathiyanarayanan et al. (2013).

GC-MS analysis showed the presence of non sugar components in bacterial exopolymer (Sutherland et al., 1977). The brominated compound 1,1,3,1"-Terphenyl,3,3,5,5-tetrabromo-5-(3,5-dibromophenyl) is found in bacterial exopolymeric substance (Figure 5). This was considered as an important characteristic feature of exopolymer. Lowell (1966) and Isnanetyo et al. (2003) also reported that the brominated secondary metabolites from the marine organisms exhibit a potent cytotoxic property. Lowell (1966) observed a low molecular weight brominated compound named 2,3,4,tribromo,5(5 hydroxy-2',2',dibromophenyl)pyrrole in *Thalassia* sp. 3,3,35,5'-tetra bromo-2-2 dihenyldiol was isolated from *Pseudoalteromonas phenolica* by Isnanetyo et al. (2003) which has strong antibacterial activity against methicillin resistant *Staphylococci aureus*.

In order to survive under stressful conditions, some of the marine bacteria synthesize modified proteins, fatty acids and sterols, bounded on their cell wall or incorporated with their exocellular substances (Weber et al., 1996). Polyunsaturated fatty acid synthesis by the marine bacteria is one of the adventitious characters to survive in an extreme habitat. Yano et al. (1997) also observed the presence of polyunsaturated fatty acid (PUFA) such as Docosahexanoic acid and eicosapentanoic acid in the cell wall of marine bacteria *Vibrio marinus* and sometimes are likely to be included with the exopolymeric substances. These polyunsaturated fatty acids are essential in maintaining the fluidity condition under high pressure and adaptation to saline and extreme temperature. These polyunsaturated fatty acids also act as a defense compound having antimicrobial activity (Hayashida et al., 2008). Meanwhile, saturated fatty acids namely Docosanoic acid, 1,2,3-propanetriyl ester and Pentanoic acid, 5-hrdroxy,2,4-di-t-butylphenyl esters were found in the exopolymer. Russell (1989) also noted that there is the presence of saturated fatty acid in *Marinococcus halophilus* for maintaining membrane fluidity and nutrient transport. Docosanoic acid, 1,2,3-propanetriyl ester from plant origin named as Tribehenin, which acts as a good skin conditioning agent and emulsifying agent (Daffodil et al., 2012).

The cytotoxic effect of bacterial exopolymer was due to the presence of brominated compound. The enzyme mitochondrial oxidoreductases facilitate reduction of slightly yellow compound XTT to orange at the cell surface by transplasma membrane electron transport. The succinate dehydrogenase system of the mitochon-drial respiratory chain reduced the tetrazolium salt of XTT (2,3-Bis-(2-methoxy-4-nitro-5-sulfophenyl]-2H-tetrazolium -5-carboxyanilide salt) to soluble formazan only in living cells. The mitochondrial membranes of live cells have active enzyme system, which are inactivated shortly after cell death. The cell lines treated with bacterial exopolymer destroy the respiratory chain and inactivate the enzyme and, therefore, fail to form a soluble orange formazan by reduction of the yellow tetrazolium salt.

This work shows the potential application of bacterial exopolymer. However, biotechnological potential of the biopolymer produced by marine bacteria remains largely unexploited. The present study will give an idea on the significance and insight into the potential of the bacterial exopolymer.

Conflict of Interests

The author(s) have not declared any conflict of interests.

ACKNOWLEDGEMENTS

REFERENCES

Adriana AT, Bernadette P, Daniela I, Concetta G, Giuseppe B (2005). Antiviral and immunoregulatory effect of a novel exopolysaccharide from a marine thermotolerant *Bacillus licheniformis*. Int. Immunopharmacol. 6:8–13.

Al Nahas MO, Darwish MM, Ali AE, Amin MA (2011). Characterization of an exopolysaccharide producing marine bacterium isolate *Pseudoalteromonas sp*. Afr. J. Microbiol. Res. 5(22):3823-3831.

Annarita P, Anzelmo G, Nicolaus B (2010). Bacterial exopolysaccharides from marine habitats Production, characterization and biological activities. Mar.Drugs. 8:1779-1802.

Barsby T, Kelly MT, Gagne SM, Andersen RJ (2001). Bugorol A produced in culture by a marine *Bacillus sp*. reveals a novel template for cationic peptide antibiotics. Org. Lett, 3:437- 440.

Bragadeeswaran S, Jeevapriya R, Prabhu K, Sophia RS, Priyadharsini S, Balasubramanian T (2011). Exopolysaccharide production by *Bacillus cereus* GU812900, a fouling marine bacterium. Afr. J. Microbiol. Res.5 (24):4124-4132.

Braissant O, Decho AW, Dupraz C, Glunk K, Przekop M, Visscher PT(2007). Exopolymeric substances of sulfate-reducing bacteria Interactions with calcium at alkaline pH and implication for formation of carbonate minerals. Geobiology, 1-11.

Bruckner J (1955). Estimation of monosaccharides by the orcinol–sulphuric acid reaction Biochem J. 60(2):200–205.

Castresana J (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17(4):540-52.

Cerning J, Renard CMCG, Thibault, JF, Bouillanne C, Landon M, Desmazeaud M, Topisirovic L (1994). Carbon source requirements for exopolysaccharide production by *Lactobacillus casei* CG11 and partial structure analysis of the polymer. Appl. Environ. Microbiol. 60:3914–3919.

Chen W, Zhao Z, Chen SF, Li YQ (2008).Optimization for the production of exopolysaccharide from Fomes formentarius in submerge

culture and its antitumor effect in vitro. Bioresour. Technol. 99:3187–3194.

Daffodil ED, Uthayakumari FK, Mohan VR (2012). GC-MS determination of bioactive compounds of *curculigo orchioides gaertn.* Science Research Reporter 2(3):198-201.

Danesi R, DeBraud F, Fogli S, De Pas TM, Di Paolo A, Curigliano G, Del Tacca M (2003). Pharmacogenetics of anticancer drug sensitivity in non-small cell lung cancer. Pharmacol. Rev. 55: 57-103.

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res.1:36.

Donlan RM, Costerton JW (2002). Biofilms;survival mechanisms of clinically relevant microorganisms. Clin.Microbiol.Rev. 15:167-193.

Gancel F, Novel G (1994). Exopolysaccharide production by *Streptococcus salivarius* ssp. thermophilus cultures: 1. Conditions of production. J. Dairy Sci. 77:685–688.

Hayashida Soiza G, Uchida A, Mori N, Kuwahara Y, Ishida Y (2008). Purification and characterization of antibacterial substances produced by a marine bacterium *Pseudoalteromonas haloplanktis* strain. J. Appl. Microbiol. 105:1672–1677.

Holt JG, Krieg NR, Sneath PHA, Staley JT, Williams ST (1994). Bergey's Manual of Determinative Bacteriology, ninth ed., Williamsons and Wilkins, Balitomore. 2:559-560.

Inmaculada LI, Mata JA, Tallon R, Bressollier P, Urdaci MC, Quesada E, Bejar V (2010). Characterization of the Exopolysaccharide Produced by *Salipiger mucosus* A3T, a Halophilic Species Belonging to the *Alphaproteobacteria*, Isolated on the Spanish Mediterranean Seaboard. Mar. Drugs.8(8):2240–2251.

Isnanetyo A, Kamei Y (2003). MC21-A, a bactericidal antibiotic produced by a new marine bacterium,*Pseudoalteromonas phenolica sp.* nov. O-BC30T against methicillin-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. 47:480-488.

Jean-Marc N, Yves VP, Lydia H, Rupert de W (1990). Compilation of small ribosomal subunit RNA sequences Nucl. Acids Res.18:2237-2317.

Jensen PR, Fenical W (1994). Strategies for the discovery of secondary metabolites from marine bacteria.Ecological perspectives. Annu. Rev. Microbiol.48:559-584.

Khuri FR, Herbst RS, Fossells FV (2001). Emerging therapies in non-small-cell lung cancer. Ann Oncol .12:739-44.

Kim MJ, Seo HN, Hwang TS, Lee SH, Park DH (2008). Characterization of exopolysaccharide (EPS) produced by *Weissella hellenica Skkimchi* isolated from kimchi. J. Microbiol. 46(5):535-541.

Lijour Y, Gentric E, Deslandes E, Guezennec J (1994). Estimation of the sulfate content of hydrothermal vent bacterial polysaccharides by fourier transform infrared spectroscopy.Anal. Biochem. 220: 244–248.

Lowell FM (1966). The structure of a bromine rich antibiotic. J. Amer. Chem. Soc. 88: 4510-4511.

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951). Protein measurement with the Folin phenol reagent. J. Biol. Chem. 193:265-275.

Lungmann P, Choorit W, Prasertsan P (2007). Physio-chemical and biological properties of partially purified exopolymers from newly isolated halophilic bacterial strain SM5 Songklanakarin J. Sci. Technol. 29(6): 1571-1582.

Mancuso Nichols CA, Garon S, Bowman JP, Raguenes G, Guézennec J (2004). Production of exopolysaccharides by Antarctic marine bacterial isolates. J Appl Microbiol. 96(5):1057-66.

Mata JA, Bejar V, Bressollier P, Tallon R, Urdaci MC, Quesada E, Inmaculada Llamas I (2008). Characterization of exopolysaccharides produced by three moderately halophilic bacteria belonging to the family Alteromonadaceae. J. Appl. Microbiol. 105:521–528.

Ng TK, Hu WS (1989). Adherenceof emulsanproducing *Acinetobacter calcoaceticus* to hydrophobic liquids. Appl. Microbiol. biotechnol. 31:480-485.

Read RR, Costerton JW (1987). Purification and characterization of adhesive exopolysaccharides from *Pseudomonas putida* and *Pseudomonas fluorescens.* Can. J. Microbiol. 33:1080-1090.

Reichenbach H (2001). *Myxobacteria* producers of novel bioactive substances. J. Ind. Microbiol. Biotechnol. 27:149-156.

Roehm NW, Rodgers GH, Hatfield SM, Glasebrook AL (1991). An improved colorimetric assay for cell proliferation and viability utilizing the tetrazolium salt XTT. J. Immunol. Methods 142:257-65.

Russell NJ (1989). Adaptive modifications in membranes of halotolerant and halophilic microorganisms. J. Bioenerg. Biomembr. 21:93–113.

Sathiyanarayanan G, Seghal Kiran G, Joseph Selvin (2013). Synthesis of silver nanoparticles by polysaccharide bioflocculant produced from marine *Bacillus subtilis* MSBN17, Colloids Surf: B 102:13– 20.

Stevens MG, Olsen SC (1993). Comparative analysis of using MTT and XTT in colorimetric assays for quantitating bovine neutrophil bactericidal activity. J. Immunol. Methods. 157:225-31.

Sung HK, Hyun SL, Shin HP, Hong KL (2000). Optimal Conditions for the Production of Exopolysaccharide by Marine Microorganism *Hahella chejuensis*. Biotechnol. Bioprocess Eng.5:181-185.

Sutherland IW (1977). Microbial exopolysaccharide synthesis. In Extracellular Microbial Polysaccharides, Sanford P A and A Laskin,Am. Chem. Soc 40–57 pp.

Vijayabaskar P, Babinastarlin S, Shankar T, Sivakumar T, Anandapandian KTK (2011). Quantification and Characterization of Exopolysaccharides from *Bacillus subtilis* (MTCC 121). Adv. Biol. Res. 5(2):71-76.

Vincent P, Pignet P, Talmont F, Boui L, Foumet B, Guezennec J, Jeanthon C, Prieur D (1994). Production and characterization of an exopolysaccharide excreted by a deep-sea hydrothermal vent bacterium isolated from polycheate annelid *Alvinella pompejana.* Appl. Environ. Microbiol. 60:4134-4141.

Weber FJ, DeBont JAM (1996). Adaptation mechanisms of microorganisms to the toxic effects of organic solvents on membranes. Biochem. Biophys. Acta Rev. Biomembr. 1286:225–245.

West TP, Strohfus B (1998). Effect of complex nitrogen sources upon gellan production by *Sphingomonas paucimobilis.* Microbios 94:145-152.

Yano Y, Nakayama A, Yoshida K (1997). Distribution of polyunsaturated fatty acids in bacteria present in intestines of deep-sea fish and shallow-sea poikilothermic animals. Appl. Environ. Microbiol. 63:2572–2577.

Full Length Research Paper

Amylase production under solid state fermentation by a bacterial isolate W74

Kindu Nibret Tsegaye[1]* and Amare Gessesse[2]

[1]Natural Science department, Biology Unit, Gondar College of Teachers' education, Gondar, Ethiopia.
[2]Biotechnology Unit, Faculty of Science, Addis Ababa University, Addis Ababa, Ethiopia.

This study was concerned with the screening of a suitable isolate and optimization of cultural conditions for the biosynthesis of thermostable amylase under solid state fermentation (SSF). Twenty seven isolates were screened for amylase production out of which one isolate designated as W74 showed maximal amylase activity at 70°C and a pH of 6.5 and selected for further optimization of cultural conditions under SSF. Among the different carbon and nitrogen sources supplemented to wheat bran, starch (96.7 U/g) and casein (107.3 U/g) enhanced maximum amylase production. Addition of exogenous glucose repressed secretion of amylase, demonstrating that a classical glucose effect was operative in this organism. Cultural optimization was undertaken to evaluate the effect of main process parameters as incubation period (144 h), moisture (66.7%), inoculum size (40%), and initial medium pH (6.5) on enzyme production. The enzyme was optimally active at 70°C and in pH range of 5.5-6.5.

Key words: Thermostable amylase, solid state fermentation, wheat bran, enzyme

INTRODUCTION

Enzymatic hydrolysis of starch is carried out under temperatures up to 100°C, normal pressure, and pH of medium around 6.0 to 8.0. However, enzymes are relatively expensive and above all thermally unstable at higher temperatures as reviewed in Hossain et al. (2006). Attempts are now being made to find enzymes from thermophilic microorganisms. Extreme environments (high temperature and acidic environments) harbor a wide range of acido-philic hyperthermophilic organisms including members of both bacteria and archaea prokaryotic subdivisions (Worthington et al., 2003). These properties imply extremely important industrial and biotechnological imply-cations due to the fact that enzymes from such microorganisms can be employed for use in harsh industrial conditions where their specific catalytic activity is retained (Haki and Rakshit, 2003). A new strain of Bacillus sp.

I-3 was isolated from natural soil samples by Soni et al. (2005) and the crude α-amylase extract showed maximum activity at 70^0C, pH 7. It has been reported by Mamo et al. (1999) that a thermostable amylase producing microbe, *Bacillus* sp. WN11, was isolated from Wondo Genet hot spring. Similarly, Haki and Rakshit (2004) isolated bacterial colonies from Ethiopian hyper-thermal springs at Arbaminch, Awassa, Nazreth, Shalla and Abijata, Wendo Genet and Yirgalem. The thermo-stability experiments showed that more thermotolerant enzymes were isolated from Shalla and Abijata, followed by Awassa, where the temperatures of the water were also the highest. The bacterial colonies were identified as *B. stearothermophilus* and *B. licheniformis* by Haki and Rakshit (2004). Furthermore, Muluye Teka (2006) has isolated a *Bacillus* sp. from Lake Chitu and found amy-lases with optimum temperature of 80^0C. The present study was aimed to isolate and screen bacterial species from hot spring soil samples for the production of thermostable amylases under SSF.

EXPERIMENTAL METHODS

Bacterial Isolation

The bacterial strains used in this study were isolated by directly inoculating 1 g of soil samples from Gendysony thermal spring located in Arbaminch area, Ethiopia, into 10 g solid substrate (wheat bran) containing 0.5 g soluble starch. The inoculated bran was statically incubated at 37°C. After 5 days of incubation a sample was taken, serially diluted and spread plated on to starch agar containing 0.5% soluble starch as used by Elnasser et al. (2007). Screening of isolates for amylolytic activity was carried out by growing the organisms on starch agar plates containing 0.5 %(w/v) starch and subsequently staining with iodine solution (1% I_2 (w/v) in 2% (w/v) KI). The presence of a halo around the colony was indica-tive of amylolytic activity. The composition of starch agar used in the study was as follows: starch (0.5%), bacteriological peptone (0.2%), $MgSO_4$ (0.02%), $CaCl_2$ (0.02%), K_2HPO_4 (0.1%) and agar 1.5%. All the materials and reagents used in this study were obtained from microbiology laboratory at Addis Ababa University, Science Faculty.

Screening for thermostable bacterial amylases

The isolated strains were further screened for their ability to produce thermostable amylases, with assay temperatures ranging from 50 to 90°C at 10°C intervals and incubation period of 10 min. The individual isolates were re-inoculated into solid media and after 5 days of incubation, enzyme was extracted with 100 mL distilled water for assaying. The isolate with the best enzyme activity at 70°C was selected and taken for further investigations.

Inoculum preparation

For inoculum preparation, 250 mL Erlenmeyer flasks containing 50 mL of starch broth were inoculated with a loop full of cells from a 24

h slant and kept in a rotary shaker (120 rpm) at room temperature. After 24 h of incubation, 3 mL of culture were used as the inoculum. By serial dilution and plating, the number of viable colonies in the inoculum was determined.

Enzyme production in SSF

The SSF process was carried out in 250 mL Erlenmeyer flasks using 10 g of wheat bran solid substrate (Table 4). After proper agitation of the substrate, it was autoclaved at 121°C for 15 min, allowed to cool to room temperature, and inoculated with 3 mL of 24 h old culture. Substrate moisture ratio was adjusted to 1:2. Subsequently, incubation was carried out statically at 37°C for five days. The SSF media flasks were gently shaken after every 24 h for uniform mixing up of substrate and inoculum.

Enzyme extraction

The extracellular enzymes from the fermented bacterial bran were extracted with distilled water (100mL) after agitated on a rotary shaker at 120 rpm for 30 min. The content was filtered and squeezed out through a cotton cloth. The filtrate was centrifuged at 10,000 × g for 10 min to separate small particles, cells, and spores. The brown, clear supernatant was used in enzyme assay as the crude enzyme.

Amylase activity assay

Amylase activity was determined by the procedure of Anto et al. (2006) using wheat starch, gelatinized on a heater, as a substrate. The reaction mixture containing 0.9 mL of 1% substrate in 0.01 M KH_2PO_4/K_2HPO_4 buffer, pH 6.5, and 0.1 mL of enzyme extract was incubated for 10 min at 50°C. The reaction was stopped by adding 2 mL of 3, 5-dinitrosalicylic acid solution (DNS) followed by heating in a boiling water bath for 5 min and cooling to room temperature. The absorbance of each solution containing the brown reduction product was measured at 540 nm. Enzyme assay was performed in triplicates and the average was calculated. One unit (U) of α-amy-lase activity was defined as the amount of enzyme that releases 1 µmol of reducing sugar as glucose per minute, under the assay conditions and expressed as U/g of dry substrate (Anto et al., 2006). The composition of the DNS reagent used in the study was as follows (g/L): phenol 2; sodium sulfite 0.5; sodium-potassium tartrate 20; NaOH 10; and dinitrosalicylic acid (DNSA) 10.

Optimization of fermentation process under SSF

The SSF of wheat bran for production of extracellular amylase was optimized by varying process conditions like time course, moisture level, inoculum size, carbon and nitrogen additives and initial medium pH. The strategy followed was to optimize each parameter, independently of the others and, subsequently, optimal conditions were employed in all experiments. In all optimization procedures, enzyme assays were performed in triplicates and the average result used in data analysis.

Time course of enzyme production

Growth media containing 10 g of wheat bran were incubated for

Table 1. Enzyme activities for the newly isolated bacterial colonies (W74 and W120) under various temperature ranges.

Temperature (°C)	Enzyme activity (%)	
	Isolate W74	Isolate W120
50	66	72
60	95	100
70	100	90
80	23	22

70°C was taken as 100 % for W74 and 60°C for W120

varying time periods (24-240 h) at 37°C with 3mL of a 24 h old culture as inoculum. Individual flasks were withdrawn with 1 day interval from 24-240 h to assess for enzyme production. Five day incubation was formerly employed for screening bacterial strains with amylolytic potential.

Initial moisture content

Substrate moisture ratio (w/v) was maintained as 1:1, 1:1.5, 1:2, 1:2.5 and 1:3, and incubated for 144 h (optimum) and inoculum level 30% at 37°C.

Effect of inoculum size

The wheat bran media were inoculated with different inoculum levels (10, 20, 30, 40 and 50% (w/v) and SSF was carried out for 144 h with 66.7% moisture content as pre-optimized growth conditions and incubated at 37°C. Inoculum size was determined by counting colony forming units using serial dilution and plating techniques. A sample (1mL) was taken from 24 h inoculating culture, serially diluted (10^{-1} -10^{-9}) and spread on to starch agar. Colonies were counted from clearly visible dilutions and log number of cells/g bran calculated.

Effect of carbon and nitrogen additives

Various carbon sources (0.05 g/g dry substrate) such as monosaccharides (xylose, glucose and fructose) and disaccharides (lactose, sucrose and maltose) were evaluated for their effect on amylase production by replacing starch in the production medium. The flasks were inoculated with 30% inoculum and incubated at 37ºC for 144 h in a 1:2 substrate-moisture ratio. The optimum carbon source was found by analyzing the results of amylase production. A starch supplemented and a control without additional carbon sources was also included for comparison. The production medium was supplemented with different nitrogen sources (0.02 g/g dry substrate) of NH_4Cl, $NaNO_3$, $(NH_4)_2SO_4$, peptone, yeast extract, casein and urea to check their effect on enzyme production.

Effect of initial pH on enzyme production

In order to investigate the effect of pH on extracellular amylase production, the pH of the starch solution (in duplicates) was adjusted with 1 N NaOH and 1 N HCl at values 4, 5, 6, 7, 8, 9, 10, 11 and 12 prior to sterilization. When wheat bran was moistened with the above starch solution, the respective final pH recorded in one duplicate was 4=5.6, 5=5.6, 6=5.9, 7=6, 8=6.2, 9=6.5, 10=6.7, 11=7.0, and 12=7.7. The final medium pH (after adding wheat bran) was taken as initial medium pH. The media were then inoculated with 30% inoculum and fermentation was carried out at 37°C for 144 h with 66.7% moisture content. The optimum initial pH of the solid substrate was determined by the standard assay procedures described above.

Effect of repeated washes on amylase extraction

The fermented bran was washed with distilled water (100 mL) for five consecutive times. Enzyme activity in the respective washes was assayed using the standard assay procedures.

Spotting starch digest on thin-layer chromatography (TLC) for identifying the type of amylase

The products liberated by the action of amylase on soluble starch were identified by spotting the starch digest and standard sugars (glucose and maltose) on a silica gel plate activated at 105°C for 10 min. About 2.5 µl of the starch digest, glucose and maltose standards were spotted on TLC plates. The plates were developed in butanol: ethanol: water (50: 30: 20) and TLC was run four times to concentrate the bands. After air drying the plates, sprayed with 30% H_2SO_4 in ethanol and dried at 105°C for 10 min.

RESULTS AND DISCUSSION

Bacterial isolation

A total of 190 bacterial colonies were taken out of which 27 showed clear halos on starch agar plates. The 27 colonies were again inoculated into wheat bran and checked for enzyme production. The crude enzyme extract for all 27 isolates was assayed at 50 and 70°C, pH = 6.5. Among the 27 isolates two of them designated as W74 and W120 showing better activity at 70°C were further analyzed for their activity at various pH and temperature ranges (Tables 1 and 2). Thus, isolate W74 showed a better activity at 70°C and was selected for further optimization experiments.

Table 2. Effect of pH on enzyme activities by isolates W74 and W120 at 50°C.

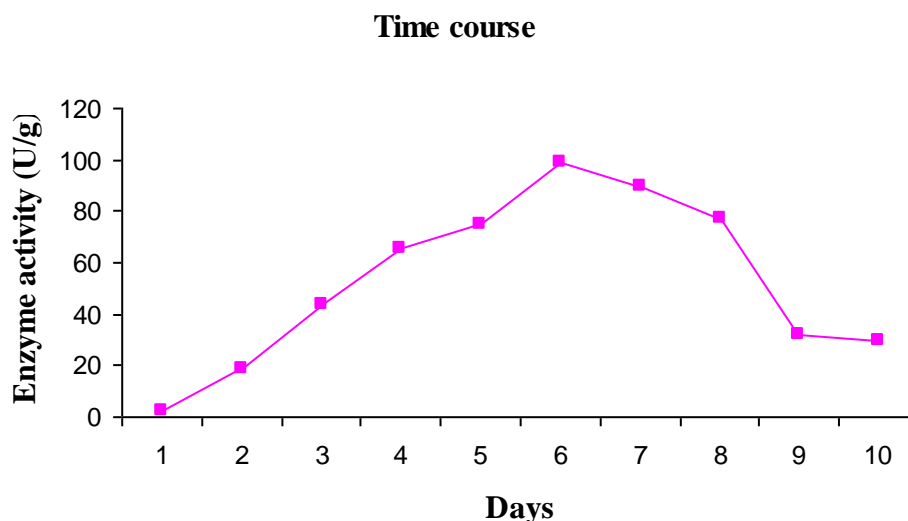| pH		3.5	4.5	5.5	6.5	7.5	8.5
Enzyme Activity (%)	Isolate W74	13	62	93	100	89	78
	Isolate W120	22	63	100	54	43	25

pH 6.5 was taken as 100 % for W74 and 5.5 for W120

Figure 1. Time course of amylase production under solid state fermentation by isolate W74.

Optimization of cultural parameters

Time course of enzyme production

The incubation time was found to affect enzyme production (Figure 1) as it was related to the growth of the organism. There was a gradual increase in enzyme production through 24 and 48 h and maximum at 144 h (98.8 U/g). This may be because the cultures might be at stationary phase as Malhotra et al. (2000) showed that enzyme production was maximal when cells entered stationary phase. The lowest enzyme production was found at 24 h (2.6 U/g) and was 29.8 U/g at the 10th day. The decline in enzyme production with prolonged incubation may be due to loss of moisture, slower growth, and lower enzyme production rates etc (Anto et al., 2006; Gangadharan et al., 2006).

Moisture levels

Moisture content changes during SSF as a result of eva-

poration and metabolic activities, therefore, adjusting moisture level can be very important. Enzyme production profiles with varying moisture levels (Figure 2) showed that SSF medium adjusted at 66.7% moisture content resulted in higher enzyme synthesis (114.1 U/g). Above 66.7%, enzyme production decreased. The effect of moisture level in SSF was also reviewed by Anto et al. (2006), Gangadharan et al. (2006) and Baysal et al. (2008). High moisture content might resulted in decreased substrate porosity, change wheat bran structure, promoting in development of stickiness, reducing gas volume, which in turn prevents oxygen penetration (Anto et al., 2006; Gangadharan et al., 2006).

Inoculum size

Inoculum size was found to be detrimental to enzyme production (Anto et al., 2006; Baysal et al., 2008; Gangadharan et al., 2006). Inoculum levels of 0.726, 0.756, 0.773, 0.786 and 0.795 (log number of cells/g bran) were assayed to determine their effect on enzyme
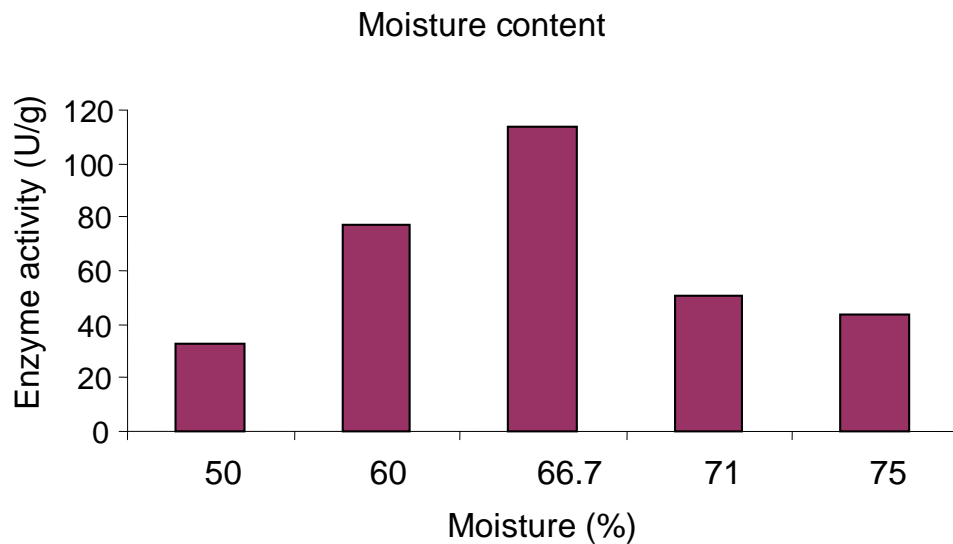
Moisture content

Figure 2. The influence of moisture content on amylase production under SSF.
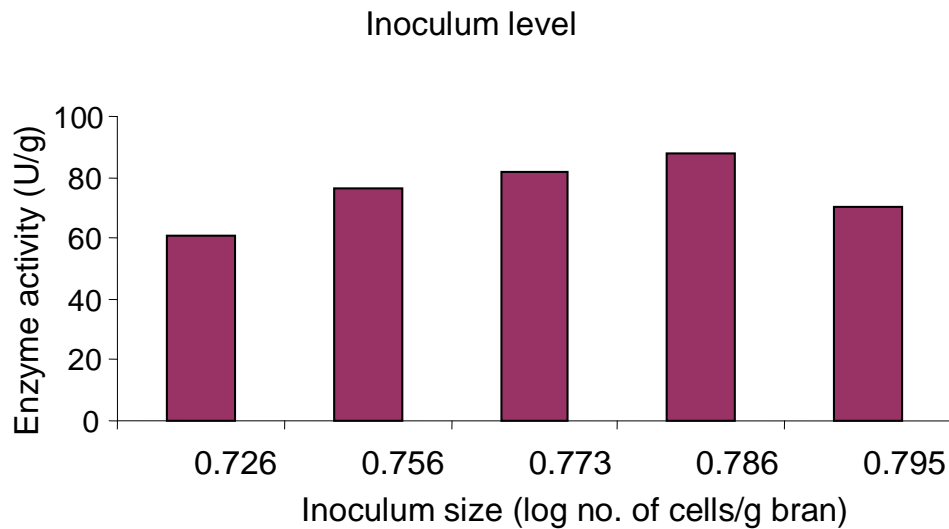
Inoculum level

Figure 3. Influence of inoculum size of α-amylase production under SSF.

production. As shown in Figure 3, the maximum enzyme production (87.7 U/g) was observed at 0.786 (optimum) and the lowest (68.9 U/g) at 0.726. Enzyme production decreased with further inoculum increments. When inoculum size was varied from 0.726-0.795, enzyme production was between 69-93% with regard to the optimum which was taken as 100%. Low inoculum size 0.726 (log number of cells/g bran) resulted in relatively lower enzyme yield. The reduction in yield at this inoculum size could be due to lower number of viable

cells for fermenting the given amount of substrate for the specified time interval. The maximum enzyme yield was obtained when 0.786 inoculum size was used. After that, there was a reduction in enzyme yield when inoculum size was increased to 0.795. This may be due to the limiting nutrients at higher inoculum size. With serial dilution and spread plating on starch agar, inoculum size (colony count) was found to be 1.8×10^7 CFU/mL, thus an inoculum size of 0.786 corresponds to 7.2×10^7 CFU/mL.
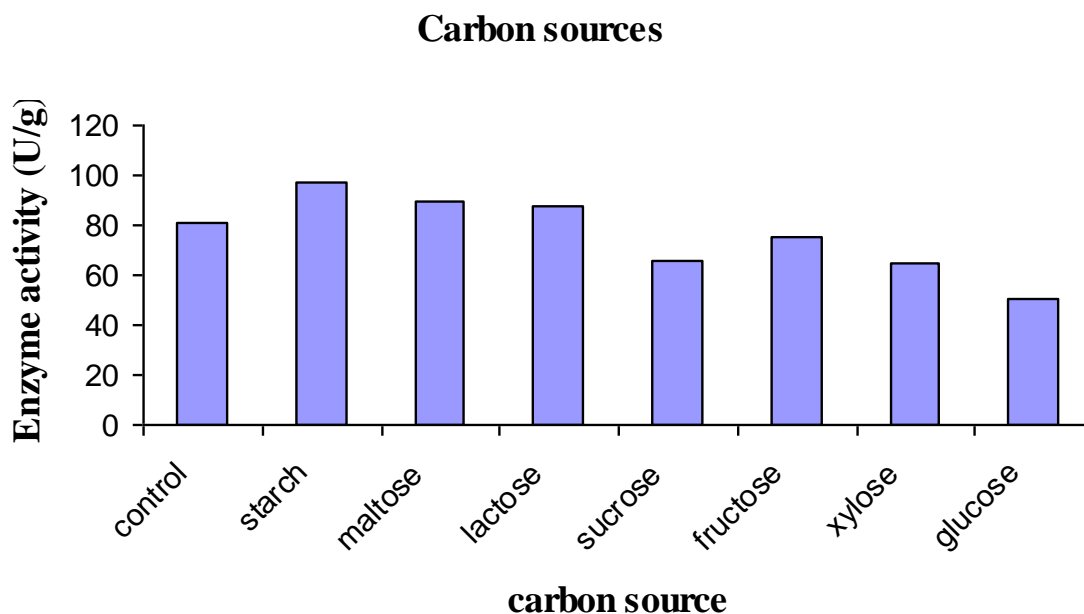
Carbon sources

Figure 4. Effects of carbon sources on amylase (U/g) production by isolate W74 under solid state fermentation using wheat bran as substrate.

Effect of supplementation of carbon sources

The highest amylase production was obtained in a medium containing starch (96.7 U/g) (Figure 4). Starch was a generally accepted nutritional component for induction of amylolytic enzymes (Mamo and Gessesse, 1999; Kiran et al., 2005; Narang and Satyanarayana, 2001; Rasooli et al., 2008). It was also observed that maltose (89.5 U/g) and lactose (87.3 U/g) favored amylase production, whereas sucrose (65.6 U/g), fructose (74.8 U/g), xylose (64.6 U/g) and glucose (50.1 U/g) gave lesser results as compared to the control which yielded 81.3 U/g. It was evident that 84% (compared to the optimum) enzyme production was recorded from the control and only 51.8% was produced when glucose was added to the fermentation medium. It has also been reported that the synthesis of amylolytic enzymes is subjected to catabolic repression by glucose (Ezeji et al. 2005; Haseltine et al. 1996; Teodoro and Martins, 2000). Presumably the same phenomenon might justify this finding.

Effect of nitrogen source on amylase production

In the investigation of the effects of various nitrogen sources on amylase production, casein (107.3 U/g) (optimum) was found to be the most promising one, followed by yeast extract (82.7 U/g) and urea (81.3 U/g)

(Figure 5). Relatively lower enzyme yields were recorded with addition of inorganic nitrogen sources and 45% reduction was recorded when $(NH_4)_2SO_4$ was used as nitrogen source. The synthesis of α-amylase was reported to be stimulated or inhibited by the type of amino acids present in the growth medium (Aguloglu et al., 2000; Park et al., 1996). In agreement with this study, organic nitrogen sources have been reported as a better inducer of amylase production than inorganic ones (Ul Qader et al., 2006; Nguyen et al., 2000).

Initial medium pH

Among the physicochemical parameters, pH of the growth medium plays an important role by inducing morphological changes in the organism and in enzyme secretion. Results showing the effect of pH on amylase production by isolate W74 in SSF of wheat bran are presented in Figure 6. The maximum activity of amylase (192.7 U/g) was observed in the fermentation medium adjusted to pH 6.7. At pH 7.7, comparatively lower enzyme production was observed. The enzyme production was much better around neutral pH ranges and only 15% was produced at pH 5.6. Microbial product formation decreases on either side of the optimum pH value (Sudharhsan et al., 2007; Ul Qader et al., 2006). In the current study, amylase production was found to be very sensitive to initial pH of the fermentation medium that is
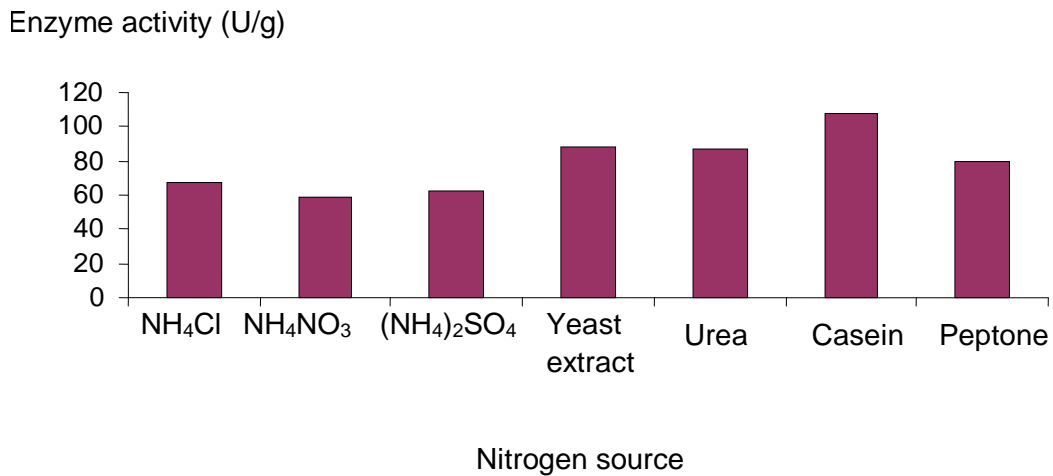
Enzyme activity (U/g)

Figure 5. Effect of nitrogen source on amylase yields under SSF by isolate W74.
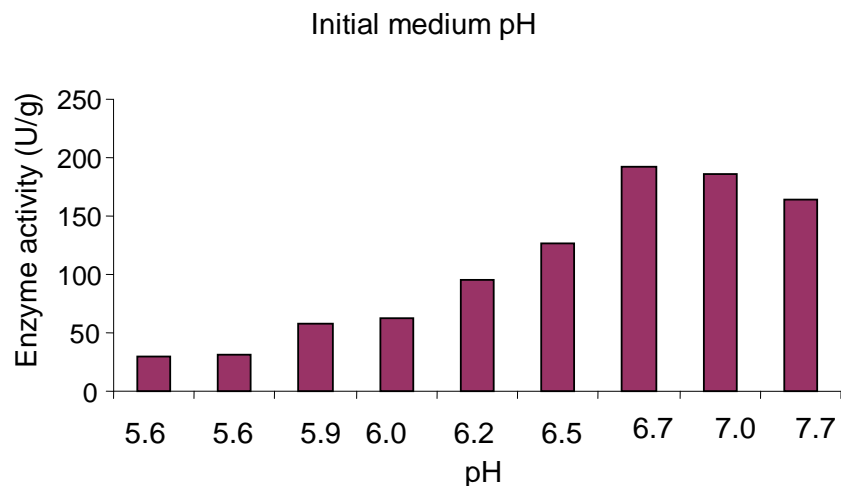
Initial medium pH

Figure 6. Effect of initial medium pH on α-amylase production by isolate W74 under SSF.

why less than 50% yield was recorded at fermentation medium adjusted to pH 6.2. It was reported by Raimbault (1998) that in SSF systems, the nature of the substrate has a strong influence on pH kinetics, due to the buffering effect of lignocellulosic materials which was similar with current study. Thus, wheat bran was examined to have a great buffering capacity.

Effect of repeated washes on amylase extraction

Extraction efficiency is critical to fully exploit the enzyme produced (Palit and Banerjee, 2001). For efficient lea-ching of the enzyme from the fermented biomass, the bacterial bran was socked for 30 minutes in five conse-cutive washes. With repeated washes, it was observed that 146.5 U/g and 14.2 U/g of enzyme extracted in the 1st and 2nd washes, respectively and only 4.5 U/g was recovered in the 3rd wash (Table 3). With the fourth and fifth washes, an almost negligible amount of enzyme was recovered. The subsequent washes did not have signi-ficant effect on extraction suggesting that most of the enzyme was leached out in the first wash. Finally, an assessment was undertaken to identify the type of amylase produced by this strain. The chromatogram indi-cated the formation of a range of oligosaccharides (data not shown) from wheat starch indicating that the enzyme was an α-amylase. Similar findings reported that α-amylases

Table 3. The effect of number of washes on α-amylase extraction from fermented bacterial bran.

Number of washes	Enzyme activity (U/g)	Percentage (%)
1st	146.5	88.3
2nd	14.2	8.6
3rd	4.5	2.7
4th	0.49	0.3
5th	0.15	0.09

Table 4. Composition (g) of solid state fermentation used for α-amylase production.

Parameter	Value (g)
Wheat bran	10
Starch (wheat)	0.5
KH_2PO_4	0.2
NaCl	0.25
$MgSO_4 . 7H_2O$	0.02
$CaCl_2 . 7H_2O$	0.02
$(NH_4)_2SO_4$	0.1

Distilled water was added to adjust the required moisture level

degrade starch in random fashion producing various maltooligosaccharide mixtures (Mamo et al., 1999; Sarikaya and Gurgun, 2000; Kiran and Chandra, 2008).

Conclusion

The results in this study indicated that isolate W74 was a potential strain for α-amylase production under solid state fermentation using wheat bran as a substrate. Its growth at neutral pH medium range and mesophilic growth temperature make isolate W74 a potential strain for future use. The extracted enzyme showed optimum activity at 70°C, pH 6.5 and was found to have a better activity between moderately acidic and neutral pH values (5.0-7.0). Therefore, research need to be undertaken to exploit the potential use of isolate W74. The utilization of wheat bran as solid substrate had a great advantage in buffering pH and its low cost could lead to large-scale production of this enzyme for industrial use in starch liquefaction.

ACKNOWLEDGEMENTS

REFERENCES

Aguloglu S, Ensari NY, Uyar F, Otludil B (2000). The effects of amino acids on production and transport of α-amylase through bacterial membranes. Starch. 52: 290-295.

Anto H, Trivedi U, Patel K (2006). Alpha-amylase production by *Bacillus cereus* MTCC 1305 using solid-state fermentation. Food Technol. Biotechnol. 44: 241-245.

Baysal Z, Uyar F, Doru M, Alkan H (2008). Production of extracellular alkaline-amylase by solid state fermentation with a newly isolated *Bacillus* sp. Prep. Biochem. Biotechnol. 38(2): 184-190.

Elnasser Z, Maraqa A, Owais W, Khraisat A (2007). Isolation and characterization of new thermophilic Bacteria in Jordan. Int. J. Microbiol. 2(3): 1-13.

Ezeji TC, Wolf A, Bahl H (2005). Isolation, characterization, and identification of *Geobacillus thermodenitrificans* HRO10, an α-amylase and a α-glucosidase producing thermophile. Can. J. Microbiol. 51: 685-693.

Gangadharan D, Sivaramakrishnan S, Nampoothiri KM, Pandey A (2006). Solid culturing of *Bacillus amyloliquefaciens* for alpha-amylase production. Food Technol. Biotechnol. 44(2): 269-274.

Haki GD, Rakshit SK (2003). Developments in industrially important thermostable enzymes: Review paper. Bioresour. Technol. 89: 17-34.

Haki GD, Rakshit SK (2004). Thermostable amylases from hyper-thermal springs of Ethiopia. Trop. Sci. 44: 57-60.

Haseltine C, Rolfsmeier M, Blum P (1996). The glucose effect and regulation of α-amylase synthesis in the hyperthermophilic archaeon

Sulfolobus solfataricus. J. Bacteriol. 178(4): 945-950.

Hossain SMZ, Haki GD, Rakshit SK (2006). Optimum production and characterization of thermostable amylolytic enzymes from *B. stearothermophilus* GRE1. Can. J. Chem. Engin. 84: 368-374.

Kiran KK, Chandra TS (2008). Production of surfactant and detergent stable, halophilic, and alkalitolerant alpha-amylase by a moderately halophilic *Bacillus* sp. strain TSCVKK. Appl. Microbiol. Biotechnol. 77: 1023-1031.

Kiran O, Comlekcioglu U, Arikan B (2005). Effects of carbon sources and various chemicals on the production of a novel amylase from a thermophilic *Bacillus* sp. K-12. Turk J. Biol. 29: 99-103

Malhotra R, Noorwez SM, Satyanarayana T (2000). Production and partial characterization of thermostable and calcium-independent α-amylase of an extreme thermophile *Bacillus thermooleovorans* NP54. Lett. Appl. Microbiol. 31: 378-384.

Mamo G, Gashe BA, Amare Gessesse (1999). A highly thermostable amylase from a newly isolated thermophilic *Bacillus* sp. WN11. J. Appl. Microbiol. 86: 557-560.

Mamo Gashaw, Amare Gessesse (1999). Effect of cultivation conditions on growth and α-amylase production by a thermophilic *Bacillus* sp. Lett. Appl. Microbiol. 29: 61-65.

Muluye Teka (2006). Amylases of potential industrial application from microbial sources. MSc thesis. Addis Ababa University.

Narang S, Satyanarayana T (2001). Thermostable α-amylase production by an extreme thermophile *Bacillus thermooleovorans*. Lett. Appl. Microbiol. 32: 31-35.

Nguyen QD, Rezessy-Szabo JM, Hoschke A (2000). Optimisation of composition of media for the production of amylolytic enzymes by *Thermomyces lanuginosus* ATCC 34626. Food Technol. Biotechnol. 38(3): 229-234.

Palit S, Banerjee R (2001). Optimization of extraction parameters for recovery of α-amylase from the fermented bran of *Bacillus circulans* GRS313. Braz. Arch. Biol. Technol. 44(1): 107-111.

Park YS, Dohjima T, Okabe M (1996). Enhanced α-amylase production in recombinant *Bacillus brevis* by fed-batch culture with amino acid control. Biotechnol. Bioengin. 49: 36-44

Raimbault M (1998). General and microbiological aspects of solid substrate fermentation: Review article. Electronic J. Biotechnol. 1 (3): 1-15

Rasooli I, Astaneh SDA, Borna H, Barchini KA (2008). A Thermostable α-amylase producing natural variant of *Bacillus* spp. isolated from soil in Iran. Am. J. Agri. Biol. Sci. 3(3): 591-596.

Sarikaya E, Gurgun G (2000). Increase of the α-amylase yield by some *Bacillus* strains. Turk J. Biol. 24: 299-308

Soni SK, Gupta JK, Goyal N (2005). A novel raw starch digesting thermostable α-amylase from *Bacillus* sp. I-3 and its use in the direct hydrolysis of raw potato starch. Enz. Microb. Technol. 37:723-734.

Sudharhsan S, Senthilkumar S, Ranjith K (2007). Physical and nutritional factors affecting the production of amylase from species of *Bacillus* isolated from spoiled food waste. Afr. J. Biotechnol. 6(4): 430-435.

Teodoro CES, Martins MLL (2000). Culture conditions for the production of thermostable amylase by *Bacillus* sp, Braz. J. Microbiol. 31: 298-302.

Ul Qader SA, Bano S, Aman A, Syed N, Azhar A (2006). Enhanced production and extracellular activity of commercially important amylolytic enzyme by a newly isolated strain of *Bacillus*. sp. AS-1. Turk J. Biochem. 31(3): 137-142.

Worthington P, Blum P, Perez-Pomares F, Elthon T (2003). Large scale cultivation of acidophilic hyperthermophiles for recovery of secreted proteins. Appl. Environ. Microbiol. 69(1): 252-257.

Full Length Research Paper

Production and partial purification of glucoamylase from *Aspergillus niger* isolated from cassava peel soil in Nigeria

Lawal, A. K.[1*], Banjoko, A. M.[2], Olatope, S. O.[1], Alebiosu, F. A.[1], Orji, F. A.[1], Suberu, Y. L.[1], Itoandon, E. E.[1], Shittu, K. A.[1], Adelaja, O. D.[1], Ojo, E.[1], Dike, E. N.[1] and Elemo, G. N.[2]

[1]Enzyme Technology Division, Department of Biotechnology, Federal Institute of Industrial Research, Oshodi, P.M.B. 21023, Ikeja, Lagos, Nigeria.
[2]Department of Food Technology, Federal Institute of Industrial Research, Oshodi, P.M.B. 21023, Ikeja, Lagos, Nigeria.

Glucoamylase is an enzyme that hydrolyses 1,4α and 1,6β-glucosidic linkages in polysaccharides yielding glucose. *Aspergillus niger* strains 1, 2 and 3 were locally isolated from cassava peel dumpsite for the production of glucoamylase enzyme. *A. niger* strains 1, 2 and 3 were screened for their hyper producing ability on potato dextrose agar using plate assay method fortified with starch agar, and showed zone of clearance of 17.0, 23.0 and 8.0 mm, respectively. The glucoamylase activity for *A. niger* strains 1 and 2 were 13 000.0 and 11 740.0, respectively. These values were however higher than the activity as obtained from the commercial enzyme with 2 500.0. Investigations on the protein (mg/ml), and specific activity (units/mg) on glucoamylase produced by *A. niger* strains 1 and 2 was 24.20, 537.19, 23.13 and 507.57, respectively. Fractionation of the enzyme ammonium sulphate (% w/v) using 60, 80 and 100% showed that the enzyme activities were 33 179.86, 47 985.86 and 19 167.65 units/ml, respectively. Protein concentrations were 16.29, 16.29 and 21.55 units/mg, respectively, while specific activities were 2 036.82, 2 945.725 and 889.45 units/mg, respectively. The production, packaging, and commercialization of glucoamylase in Nigeria will save a lot of foreign exchange earnings, and boost the economy of Nigeria.

Key words: Glucoamylase, specific activity, *Aspergillus niger*, fractionation, cassava peel.

INTRODUCTION

Glucoamylase is one of the oldest and widely used enzymes in food industry. They are microbial enzymes, present in bacteria, archaea, and fungi but never in plants and animals. Fungal glucoamylases are biotechnologically very important as they are used industrially in large amounts, and have been extensively studied during the

past 30 years (James and Lee, 1997). Glucoamylase (EC 3.2.1.3) hydrolyzes polysaccharides from the non-reducing bonds consecutively. They hydrolyze both 1, 4α and 1,6β-glucosidic linkages in polysaccharides yielding glucose only (Kumar and Satyanarayana, 2009). Glucoamylases are used mainly in the production of glucose syrup, high fructose corn syrup and in whole grain and starch hydrolysis for alcohol production. The major characteristic of glucoamylase is the saccharification of partially processed starch/dextrin to glucose, which is an essential substrate for numerous fermentation processes in a range of food and beverage industries (Lowry et al., 1951). Glucoamylase for commercial purposes has traditionally been produced employing filamentous fungi; although a diverse group of microorganisms is reported to produce glucoamylase, since they secrete large quantities of the enzyme extracellularly.

The commercially used fungal glucoamylases have certain limitations such as moderate thermostability, acidic pH requirement and slow catalytic activity that increase the process cost, consequently, the search for newer glucoamylase and protein engineering to improve pH and temperature optima leading to amelioration in catalytic efficiency of existing enzymes have been the major areas of research over the years (Marin-Navano and Polaina, 2011). The high cost of importation had over time being as a result of patent rights (Okafor, 1989) but improvement in optimal enzymatic activities has lead to this investigation. The use of locally available agro waste resources and imploring microorganisms in production is expected to reduce cost of importation and eventually improve quality of products in relevant industries. The objectives of this study were to isolate, screen for glucoamylase hyper producing strains of A. species and establish partial purification of produced glucoamylase using ammonium sulphate precipitation.

MATERIALS AND METHODS

Collection of soil sample

Soil samples from cassava processing areas were obtained from the cassava processing pilot plant site of The Federal Institute of Industrial Research, Oshodi extention at Tigbo'lu, Abeokuta, Ogun State. The soil samples were obtained aseptically and transferred to the laboratory for further analysis. The decision to isolate from this soil receiving cassava effluent is that the soil has high concentration of polysaccharide, and surviving microorganisms in the soil must have acquired genes for utilizing the polysaccharides. These genes include glucoamylase expression genes.

Isolation and Identification of Aspergillus spp. from soil samples

Using serial dilution method of isolation, 0.1 ml was plated on potato dextrose agar (fortified with 0.1% streptomycin). The molds were purified to obtain pure culture of A. niger. These relevant A. niger mould strain were then transferred on PDA slants kept in the refrigerator at 4°C. The mould isolates were identified following microbiological method of identifications through staining reagents of Lacto phenol cotton blue and observed under the microscope. Morphological and microscopic appearances of the moulds were compared with standard Atlas.

Screening for hyper-producing strain of Aspergillus niger

Spore suspension of each mould strain of A. niger was carefully introduced within 0.3 mm diameter of PDA plates fortified with 1% starch. The plates were then incubated separately at 30°C for 24 to 28 h, and flooded with Grams iodine reagent. Zones of clearance around the smeared portion was measured with the aid of meter rule and recorded against the blue black colouration background of PDA-starch agar plates. The zone of clearance produced is the measure of the ability of the strains with the highest ability of producing yield of glucoamylase.

Enzyme production medium

The culture medium used in this work for glucoamylase production contained $MgSO_4.7H_2O$ (0.417 g), $CaCl_2$ (0.417 g), $FeSO_4.7H_2O$ (0.208 g), $(NH4)_2SO_4$ (0.417 g), KH_2PO_4 (0.417 g), distilled H_2O (333.3 ml), rice bran (333.3 g), soya bean flour (100 g). The mineral water was prepared as stated above and the pH was adjusted to 4.5 using 2 N NaOH.

The above mixture was autoclaved at 121°C for 1 h. It was allowed to cool overnight in the autoclave. The culture microorganism (sporulating A. niger) in petridishes were washed using Tween 80. The inoculum was mixed thoroughly in a safety cabinet previously sterilized using ultraviolet (UV)-light with the substrate and covered with foil paper.

Glucoamylase assay

The reaction mixture consists of 1 ml of the enzyme extract; 1 ml of 1% (w/v) soluble starch solution and 0.1 ml citrate buffer (pH 4.5), incubation was at 60°C for 1 h. The reaction mixture was stopped by immersing the tubes in boiling water bath (100°C). The reducing sugars liberated were estimated by 3, 5 dinitrosalicylic (DNS) acid method (Mehta et al., 2006). The reaction mixture consisted 1 ml DNS reagent and 3 ml of starch hydrolysate in a test tube. The test tubes were covered and placed in boiling water (100°C) for colour development for 5 min, after which the tube was allowed to cool at ambient temperature. The absorbance was read at 540 nm against a blank using the spectrophotometer (Unispec 23D model). The blank was made up of 1 ml of the reagent with 3 ml distilled water. One unit of enzyme activity (U) was defined as the amount of the enzyme that liberated 1.0 μmole of maltose from starch in 1.0 μL reaction mixture under the assay conditions.

Protein determination

Protein content of the enzyme extracts were determined by following the method of Miller (1959) with bovine serum albumin as standard. 0.2ml protein extract, was measured into tubes and 0.8 ml distilled water was added to it. Distilled water was used as blank

while BSA standard curve was equally set up 10 mg/ml, 1 to 10 mg/ml, 5.0 ml of alkaline solution was added to all the tubes, mixed thoroughly and allowed to stand for 10 min, 0.5 ml of Folin- C solution was added to all the test tubes and left for 30 min after which the optical density was read at 600 nm wavelength in a spectrophotometer (T70 PG Instrument UV model). The protein concentration was estimated using values extrapolated from the standard graph of protein.

Ammonium sulfate fractionation of glucoamylase

Glucoamylase samples (200 ms^{-1}) were first brought to 20% (w/v) saturation with solid ammonium sulphate (enzyme grade) as mention by (Dixon and Webb (1964). The precipitated proteins were regimented by centrifugation for 15 min at 500 min^{-1}. The resulted pellet was dissolved again with ammonium sulphate to achieve 60, 80 and 100% (w/v) saturation; the precipitated proteins were centrifuged for 15 min at 500 min^{-1}. Both enzyme activity and protein content were determined for each separate fraction.

RESULTS AND DISCUSSION

Screening for the best strain of *Aspergillus niger*

All the mould strains were identified as *A. niger* but screening was carried to identify three (Lowry et al., 1951) most relevant strains with hyper production properties, which were then preserved on PDA slants and kept in the refrigerator at 4°C for further use (Table 1)

In addition, all the strains of *A. niger* could produce hydrolytic enzymes, however based on the results as obtained, *A. niger* strain 2 had the highest zone of clearance with 23 mm and was inoculated into the enzyme production medium for the production of glucoamylase (Table 1). Among the several growth factors particularly substrate particle size is one of the most critical parameter affecting the productivity and growth of microorganisms (Sen et al., 2009). It has been reported in literature that a low cost substrates like wheat flour, soya bean flour, wheat bran, rice straws (husk), rice bran and molasses are suitably effective for growth and enzyme production (Zadrazil and Punia, 1995).

Enzyme activity of crude glucoamylase

The activity of produced glucoamylases from rice bran and the commercial sample were determined, and from Table 2, the concentration of amylase produced was 4.45 mg/ml, 13000.0 µmol/L and 4.02 mg/ml, 11740.0 µmol/L, respectively; whereas the commercial amylase was 0.45 mg/ml, 2,500.0 µmol/L. The result thus shows a higher activity of produced glucoamylase enzymes when compared with the commercial enzyme sample (Table 2a and b). Ominyi et al. (2013) reported activity within 0.068 to

Table 1. Diameter of zones of clearance (mm) of *Aspergillus niger* strains.

Aspergillus niger strain	Diameter of zones of clearance (mm)
1	17.0
2	23.0
3	8.0

1.327 (Unit/ml) for glucoamylase enzyme from different fungal strains (Ominyi et al., 2013). Ominiyi et al. (2013) also reported the increased glucoamylase activity when mixed cultures of fungal strains were involved in the fermentation process (Ominyi et al., 2013). This synergistic increase in the activity of glucoamylase on utilization of mixed microbial cultures for fermentation has been reported also by Ueda (Oyewole and Agboola, 2011). There have been several attempts by other Indigenous Nigerian Scientists to extract glucoamylase from non-microbial sources including kilned and unkilned malted sorghum and corn. Oyewole and Agboola (2011) reported a glucoamylase activity of 2.78, 0.94, 1.22, and 0.75 (Mg/Maltose/ml/min) in malted unkilned sorghum, kilned sorghum, unkilned corn and kilned corn, respectively (Oyewole and Agboola, 2011). The general problems with plant enzymes are that their production is not sustainable for commercialization because continuous use of the grains will lead to loss or threat to the plant species (negative impact on biodiversity), again the malting process of the grains is not only labour intensive but also time-consuming.

Extraction and purification of glucoamylase

The enzyme was harvested on the 6th day using 2.5 L of 0.01 M citrate phosphate buffer (pH 4.5); 5 g of sodium benzoate was added to prevent enzyme deterioration by microbial actions. After adding the buffer and sodium benzoate to the medium, it was allowed to stand for 30 min and thereafter filtered using sieve packed with muslin cloth. The pH of the enzyme was 4.5. Ten (10) ml of the crude enzyme was centrifuged at 8 500 rpm for 20 min to get clear supernatant. The supernatant with glucoamylase activity of 13 000 U/ml and specific activity of 537.19 U/mg was used as crude enzyme solution and subjected to partial purification by ammonium sulphate precipitation in three fractions; 60, 80 and 100%. The crude enzyme was precipitated at 60, 80 and 100% saturation with specific activity of 2 036.82, 2 297.04 and 889.45 U/mg, respectively (Table 3).

Studies carried out so far have shown that *A. niger* strains

Table 2a. Activity of glucoamylase samples using dinitrosalicylic acid reagent at 540 nm.

Sample	Volume of standard (ml)	Volume of water (ml)	Optical density (O.D)	Concentration (mg/ml)	Concentration (µ/mol/L)
Blank	-	3.0	0.000	-	
STD. 1	0.6	2.4	0.360	0.6	
STD. 2	1.2	1.8	0.856	1.2	
STD. 3	1.8	1.2	1.376	1.8	
STD. 4	2.4	0.6	1.861	2.4	
STD. 5	3.0	-	2.282	3.0	
Glucoamylase 1	-	-	1.192	4.45	1 3000.0
Glucoamylase 2	-	-	1.075	4.02	1 1740.0
Commercial Enzyme	-	-	0.808	0.45	2 200.0

Aspergillus niger strains 1 and 2 were the sources of Glucoamylase 1 and 2 respectively.

Table 2b. Enzyme activity of crude glucoamylase.

Sample	Enzyme activity (unit/ml)	Protein content (mg/ml)	Specific activity (unit/mglml)
Glucoamylase 1	13000	24.20	537.19
Glucoamylase 2	11740	23.13	507.57

Aspergillus niger strains 1 and 2 were the sources of glucoamylase 1 and 2, respectively.

Table 3. Glucoamylase activity after Ammonium Sulphate precipitation.

Percentage (%) (w/v) Ammonium sulphate saturation	Enzyme activity (unit/ml)	Protein content (mg/ml)	Specific activity (unit/mg/ml)
60	33179.86	16.29	2036.82
80	47985.86	16.29	2945.725
100	19167.65	21.55	889.45

are good potential producers of glucoamylase using rice bran and other agro waste residues. This in turn shows the relevance and importance of agricultural raw materials found greatly in our local environment. This investigation also shows the control of regulatory patenting rights through proper research and development. Industries also have the potential of adopting the enzymes produced for optimal production and by saving cost in enzyme importation. The report also shows the need for improved enzyme production through analysis of better substrates and also adopting improved technology in production medium.

REFERENCES

Dixon M, Webb EC (1964). Enzymes. 2nd Edn. Longmans Publishers, London

James JA, Lee BH (1997). Glucoamylase: Microbial sources, industrial applications and molecular biology. A Rev J. F Biochem. 21:1-52.

Kumar P, Satyanarayana T (2009). Microbial glucoamylase characteristics and applications. Crit. Rev. Biotechnol., 29 (3):225-55.

Lowry OH, Rosebro ugh NJ, Farr AL, Randall RJ (1951). Protein measurement with the Folin phenol reagent. J. Biol. Chem. 193:265–275.

Marin-Navarro J, Polaina J (2011). Glucoamylases: Structural and biotechnological aspects, Appl. Microbiol. Biotechnol. 89 (5):1267-1273.

Mehta VJ, Thumar JT, Singh SP (2006). Production of alkaline protease from an alkaliphilic actinomycete. Biores. Technol., 97(14):1650-1654.

Miller GL (1959). Use of Dinitrosalicylic Acid reagent for determination of reducing sugars. Anal. Chem. 31:426-428.

Okafor N (1989). Patents and free communication in Industrial Microbiology. Industrial Microbiology, University of Ile-Ife Limited, Ile Ife, Nigeria.

Sen S, Veeranki VD, Mandal B (2009). Effect of physical parameters, carbon and nitrogen sources on the production of alkaline protease from a newly isolated *Bacillus pseudofirmus* SVB1. Ann. Microbiol. 59(3):531-538.

Zadrazil F, Punia AK (1995). Studies on the effect of particle size on solid state fermentation of sugar cane bagasse into animal feed using white rot fungi. Biores. Technol. 54:85-97.

Ominyi MC, Ogbonna JC, Nwoba EG, Nwagu KE, Ukachi R (2013). Isolation, and screening of α-amylase and Glucoamylase producing fungi and their Glucomaylase producing fungi and their application in Bioethanol, production. Int. J. Sci. and Nat. 4(1):44-50

Oyewole OI, Agboola FK (2011). Comparative studies on properties of amylases extracted from Kilned, unkilned malted Sorghum and Corn. Int. J. Biotechnol. Mol. Biol. Res. 2(9): 146-149.

Full Length Research Paper

Heterologous expression and characterization of purified partial endochitinase (*ech*-42) isolated from *Trichoderma harzianum*

Pratibha Sharma[1*], Manika Sharma[1] and Mukesh Srivastava[2]

[1]Division of Plant Pathology, Indian Agricultural Research Institute, New Delhi, India.
[2]Department of Plant Pathology, CSAU, Kanpur, India.

Chitinase gene from *Trichoderma harzianum* was cloned and hetrologously over expressed in M15 *Escherichia coli*. The recombinant protein of 42 kDa from *E. coli* was purified through Ni-NTA affinity column chromatography. The purified enzyme was active over broad range of pH (2.0 to 8.0) and temperature (10 to 60°C) with the peak activity at pH 5 (0.50 µg/ml) and 20°C with enzyme activity value (0.49 µg/ml). The purified protein fractions were tested for *in vitro* antifungal activity against different phytopathogens like *Fusarium oxysporum* f.sp. *lycopersici*, *Sclerotium rolfsii*, *Alternaria brassicae* and *Alternaria brassicicola*. Purified endochitinase isolated from *T. harzianum* caused necrotic lesions, segmentation, branching and hyphal bursting at the concentration of 200 µg ml^{-1}.

Key words: Antifungal activity, *Trichoderma harzianum*, *Fusarium oxysporum* f.sp. *lycopersici*, *Sclerotioum rolfsii*, *Alternaria brassicae* and *Alternaria brassicicola*.

INTRODUCTION

Trichoderma spp. are among the most frequently isolated soil fungi, well known for their biocontrol ability against a wide range of plant pathogenic fungi (Howell, 2003; Sharma et al., 2011). There are various modes of action associated with the ability of *Trichoderma* spp. to control plant pathogens but the best characterized and studied mechanism is mycoparasitism which involves a group of genes that encode for lytic enzyme chitinases (EC 3.2.1.14), known to play important role in digesting chitin wall of the phytopathogenic fungi. Chitin, which is a significant component in the cell walls of large groups of

fungi is made up of units of N-acetylglucosamine, linked together by 1,4-β-glycosidic bonds. Chitinases being chitin-degrading enzymes hydrolyze the β-1,4-glycosidic bonds between the N-acetyl glucosamine residues of chitin. There are many chitinase-producing organisms including bacteria (Ningthoujam et al., 2009), insects (Merzendorfer and Zimoch, 2003), plants (Salami et al., 2008), fungi (Rattanakit et al., 2007) and vertebrates (Tunc et al., 2008). Chitinases are classified into two families, family 18 and 19, based on the amino acid sequence of their catalytic domains. While family 18 includes

chitinases of viruses, fungi, bacteria, animals and some plants, family 19 includes all chitinases from plant origins and *Streptomyces griseus* (Patil et al., 2000; Dahiya et al., 2006). Chitinase secretion is one of the mechanisms used by *Trichoderma* to inhibit the growth of other fungi. *Trichoderma* chitinases belong to the glycosyl hydrolase family 18 and can be further grouped into class III and class V. Many chitinase genes from *Trichoderma* have been studied, including class III chitinases, such as *cht33* of *T. harzianum*, *cht33* of *T. atroviride* and *cht1* of *T. virens* and class V chitinase, such as *ech1*, *ech2* and *ech3* of *T. virens* (Kim et al., 2002; Markovich and Kononova, 2003; Duo-Chuan, 2006). Many biological control agents in the last few years were being tested and are commercially available in market. However, there is still considerable interest in finding more efficient strains, which differ considerably with respect to their biocontrol effectiveness. The main objective of the study was the heterologous expression and partially purified *ech*-42 endochitinase and *in vitro* antifungal property of purified protein against different phtyo-pathogenic fungi.

MATERIALS AND METHODS

Collection of *Trichoderma* spp. under test

Fungal cultures of *T. harzianum* were taken for isolation, cloning and characterization of gene coding endochitinase in Trichoderma. *T. harzianum* strain Th3 (ITCC: 5593), was specifically taken for this study from biocontrol laboratory, division of Plant pathology, IARI, New Delhi. The plasmid pQE-30 (Qiagen) was used as expression vector, and *E. coli* strain M15 was used as an expression host. The experimental *T. harzianum* was multiplied on potato dextrose agar media, with the combination of peeled potato: 250 g, dextrose: 20 g, agar: 15 g and distilled water: 1000 ml.

Cloning and sequencing of endochitinase gene

DNA sequence of 1,476 base encoding endochitinase was successfully cloned with accession no JN798187. The deduced amino acid sequence of endochitinase of *T. harzianum* was seen homologous to other *Trichoderma* species (82-97%) identity (Sharma et al., 2012). The nucleotide sequence having 1 to 268 bp open reading frame that has high homology with other reported Chit42 belonging to the *Trichoderma* spp. The amino acid N-terminal sequence showed a putative signal peptide for the possible secretion of the protein. The amplicon was then digested with restriction enzymes and was ligated into pre-digested pQE-30 expression vector. The recombinant vector was transformed into chemical competent cells of *E. coli* M15 by heat shock method (Froger and Hall, 2007). The transformants were selected on Luria-Bertani (LB) agar supplemented with ampicillin (100 µg.ml^{-1}). The positive clones were screened by colony PCR. The recombinant plasmids were extracted by plasmid extraction kit (Qiagen) and selection was carried out by restriction analyses and sequencing.

Expression in *Escherichia coli* and purification

The *E. coli* strain M15 harboring the pQE-Chi42 vector was grown at 37°C. When the OD 600 increased from 0.8 to 1.0, isopropyl-β-D-thiogalactoside (IPTG) inducer was added to a final concentration

of 0.1 mM. The culture was further incubated for 6 h. Cells were harvested by centrifugation and stored at -20°C. Previously stored cells at -20°C were transferred to ice and kept on ice for 30 min to thaw. The cells were resuspended in lysis buffer (50 mM Na-phosphate buffer pH 8.0, 300 mM NaCl, and 10 mM imidazole). Lysozyme was added at a final concentration of 1.0 mg ml^{-1} to facilitate the lysis. The cell lysis was performed mechanically using sonicator equipped with sonication microtip, using six 10 s bursts at 200 to 300 Watt, with a 10 s cooling period between each burst. After lysis, cell debris were removed by centrifugation at 20 000 *g*. Total soluble proteins were applied to Ni-NTA matrix column (Qiagen) and Chi42 was purified following the instruction of the manufacturer (Qiagen). Purity of the Chi42 was further confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, 12.5%) for analysis of purified protein fractions.

Enzymatic assay for chitinase

Chitinase activity is determined colorimetrically by detecting the amount of N-acetylglucosamine (GlcNac) released from the colloiddal chitin substrate. Colloidal chitin (Sigma Chemicals Co., USA) was used as a substrate (Wen et al., 2005). 0.3 ml of 1% colloidal chitin in acetate buffer (50 mM, pH 6.0) with 1 ml of enzyme were mixed and incubated at 30°C for 30 min. The hydrolysis reaction was terminated by adding 0.6 ml of dinitrosalicylic acid (DNS) reagent. The mixture was kept in a boiling water bath for 15 min, chilled and centrifuged to remove the insoluble chitin. The resulting adduct was measured in UV double beam spectrophotometer at 540 nm (Miller, 1959). One unit of enzyme activity was defined as the amount of enzyme that catalyzed the release of 1 mol of N-acetyl D-glucosamine per ml in minutes.

Effect of temperature, pH on purified enzyme activity

Chitinase activity was assayed at different pH values (pH 2.0, 3.0, 4.0, 5.0, 6.0, 7.0 and 8.0) using buffers, such as citrate-phosphate buffer (100 mM, pH, 2.5 to 7.0), sodium phosphate buffer (100 mM, pH, 7.0-8.0), and glycine-NaOH buffer (100 mM, pH,8.5 to 10.0). The purified enzyme was incubated at a range of temperatures (10, 20, 30 40, 50 and 60°C) to check for the enzyme activity.

Antifungal activity assay

Cylinder plate method (Jhonson and Curl, 1972) was used to make wells in the medium. Plug colony of *F. oxysporum* f. sp. *lycopersici* , *S. rolfsii*, *A. brassicae* and *A. brassicicola* were grown on petri dishes containing 1.5% potato dextrose agar (PDA). After three days of growth when colony diameter was 4 to 5 cm, wells were filled with 50 µL of enzyme solution with different concentration of purified enzyme ranging from (50,100, 150, 200 µg ml^{-1}) in triplicates. Plates were then observed after 16 h. Abnormal growth of hyphae and morphological changes were noted through light microscopy.

RESULTS

Purification of endochitinase

Chitinase cDNA was cloned in a prokaryotic expression vector named pQE-30. The cloning strategy was designed such that the protein containing an additional N-terminal methionine residue and C-terminal 6x-His Tag
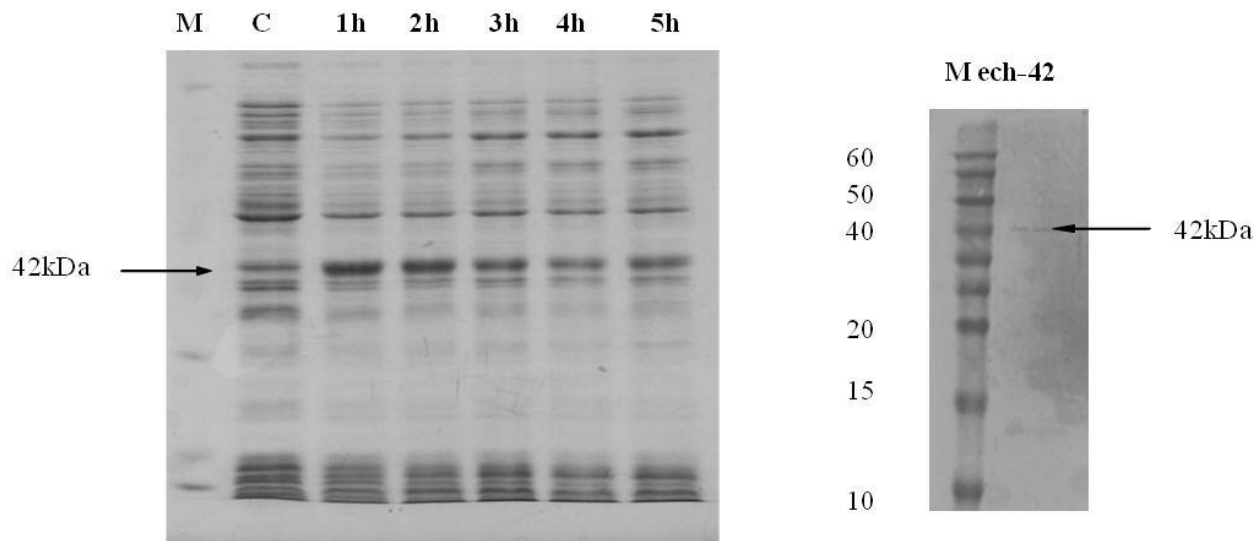
Figure 1. Hetrologous expression of endochitinase in *E. coli* after 5 h of induction with IPTG and enzyme purification through Ni-NTA column.

would be produced. Expression of the active chitinase was optimized by inducing it with IPTG inducer at different time of induction. After SDS-PAGE, the chitinase was regenerated by the removal of SDS with purified Triton X-100. After purification with Ni-NTA coloumn to bind to 6x His Tag the recombinant protein was found to be 42 kDa on SDS-PAGE as shown in Figure 1. The purified recombinant protein fractions extracted from pQE-30 with low amounts having protein (mg/mL) 1.01, 0.94, 1.20 respectively.

Effect of pH

The effect of pHs and stability for chitinase based on enzyme activity were examined at 28°C by varying pHs of the reaction mixture (pH 2 to 8) using different buffers. The enzyme was found most active between pH 3.0 to 6.0 with enzyme activity (0.30 µg/ml at pH 3.0, 0.37 µg/ml at p H 4.0, 0.57 µg/ml at p H 5.0 and 0.50 µg/ml at p H 6.0 respectively. Beyond this there is indeed loss of enzyme stability it found relatively stable at pH 7.0 and p H 8.0 with the enzyme activity (0.39 µg/ml and 0.30 µg/ml respectively). Chitinases, including the one under study showed a pH optimum in the acidic range (Figure 2).

Effect of temperature

When the enzyme was kept at various temperatures for 30 min in an acetate buffer (pH 5.0) the chitinase activity was most active at 20°C with the enzyme activity (0.49 µg/ml). The activity starts decreasing from 30°C (0.43 µg/ml) to 40°C (0.27 µg/ml). Above 40°C, the activity

decreased further at 50°C (0.18 µg/ml) and was reduced to 0.06 µg/ml at 60°C (Figure 3).

Inhibition of fungal growth

Inhibition zone of 42 kDa endochitinase isolated from *T. harzianum* against different phytopathogens like *F. oxysporum f.sp. lycopersici*, *S. rolfsii*, *A. brassicae* and *A. brassicicola* was formed at the concentration of 200 µg ml^{-1}. No inhibition zone was found between 50 to 100 µg ml^{-1}. Microscopic examination revealed that enzyme caused necrotic lesions, branching, hyphal bursting at the concentration of 200 µg ml^{-1} (Figures 4 to 7).

DISCUSSION

Fungi of genus *Trichoderma* has been long recognized for their ability to act as a biocontrol agents against plant pathogens (Harman, 2006). Most of these studies are based on characterization of genes and a few involved in purification of proteins. In the present study, an attempt was made to demonstrate *in vitro* over expression of the product of the gene coding endochitinase *ech*-42 in *E. coli* which is a part of the chitinolytic enzyme system of fungi and mechanism of biocontrol. The investigation leads us to separate a 42 kDa protein of *T. harzianum* on SDS-PAGE which was later purified to homogeneity as also been reported by Harighi et al. (2007). The activity of *ech*-42 was examined at different temperatures (20 to 60°C) with an optimum of 0.49 µg/ml at 20°C. The enzyme seems to retain activity at 60°C which supports the finding of Harighi et al. (2007) who reported the heat
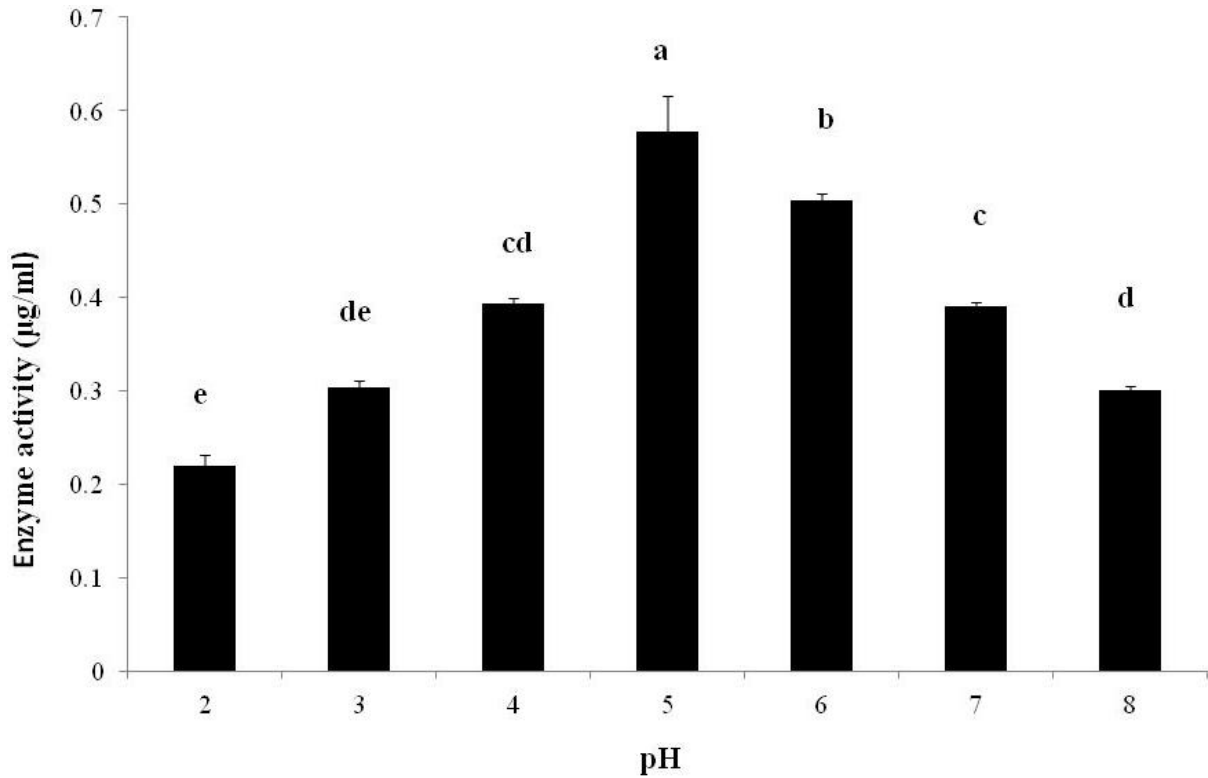
Figure 2. Influence of p H (2 to 8) on chitinase* activity. The values are the means ± S.E., for each p H , in each column followed by the same letter are significantly Different (P ≤ 0.05) from each other according to Duncan's Multiple Range (DMR) test.
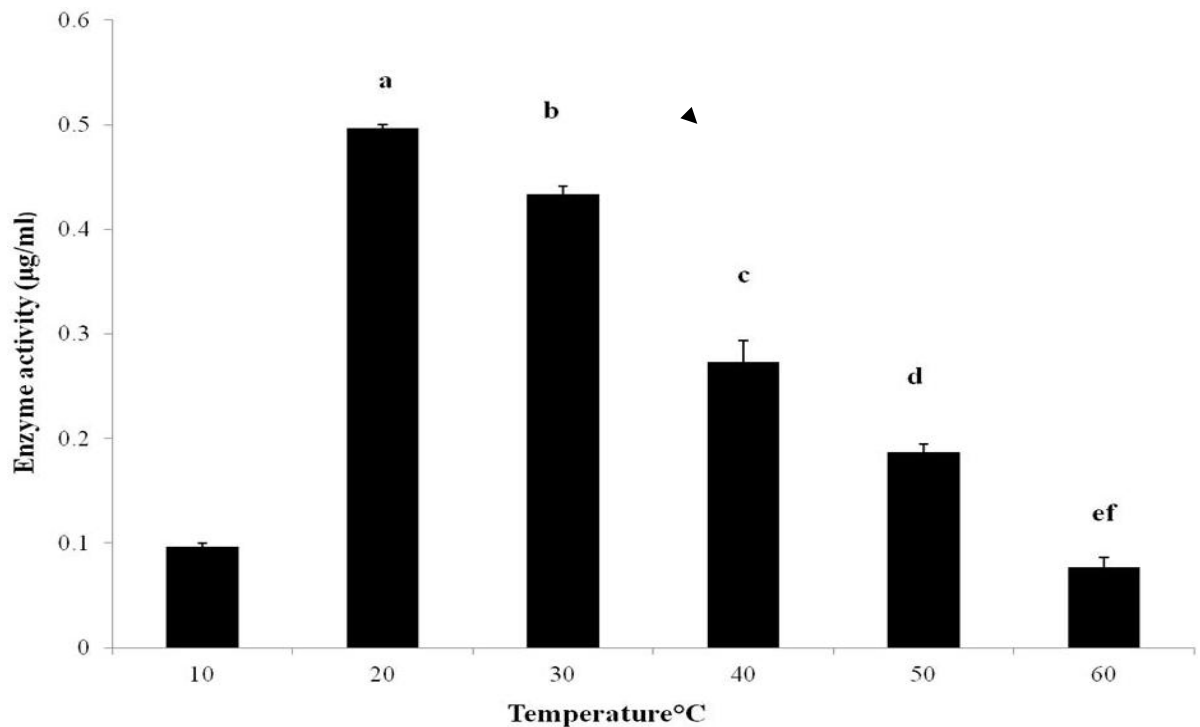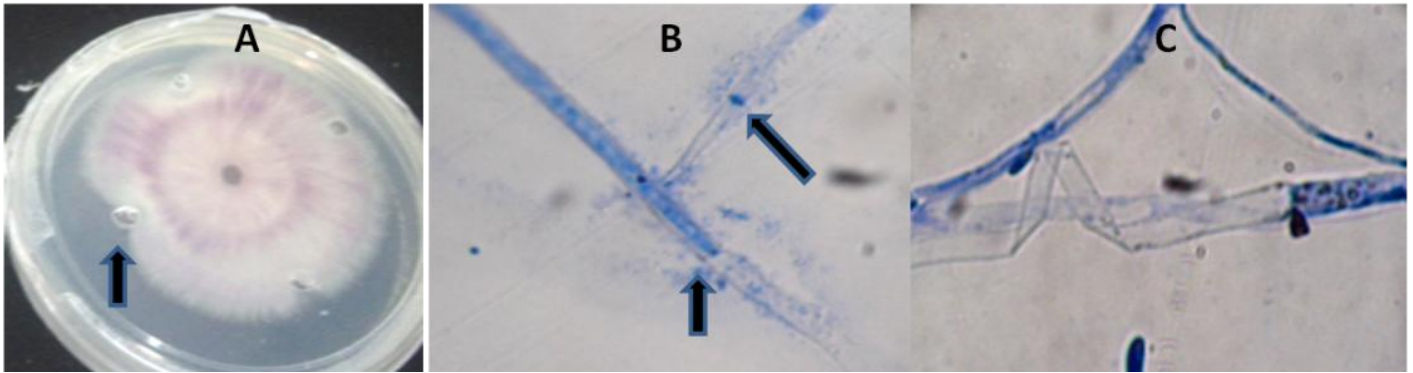

Figure 3. Influence of temperature 10 to 60°C on chitinase activity. The values are the means ± S.E. for each value in each column followed by the same letter are significantly different (P ≤ 0.05) from each other according to Duncan's multiple range (DMR) test.

Figure 4. (A) Effect on growth of *Fusarium oxysporum f.sp. lycopersici* in response to different concentrations of endochitinase. Wells contain 50 µl of enzyme solution in µgml^{-1} from top in clockwise direction (50, 100, 150 and 200). **(B)** Cellular bursting was observed at 200 µgml^{-1} concentration. **(C)** Lysis of the hypha and loss of cellular components at 200 µgml^{-1}.
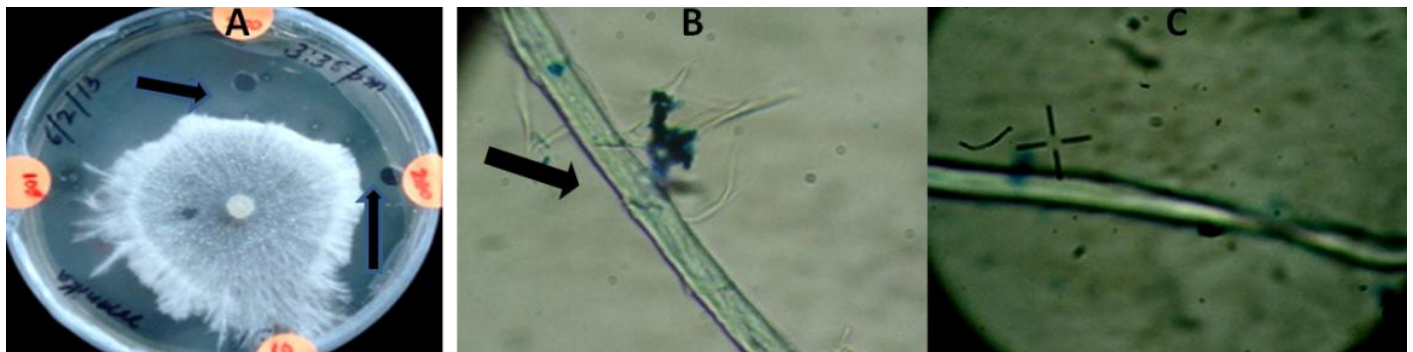

Figure 5. (A) Effect on growth of *Sclerotium rolfsii* in response to different concentrations of endochitinase. Wells contain 50 µl of enzyme solution in µgml^{-1} from top in clockwise direction (200, 100, 150 and 50). (B) Enzyme disturbance seen in the form of hyphal swelling at 200 µgml^{-1}. (C) Enzyme disturbance in the form of necrotic lesion and hyphal burst is visible at enzyme concentration of 200 µgml^{-1}.
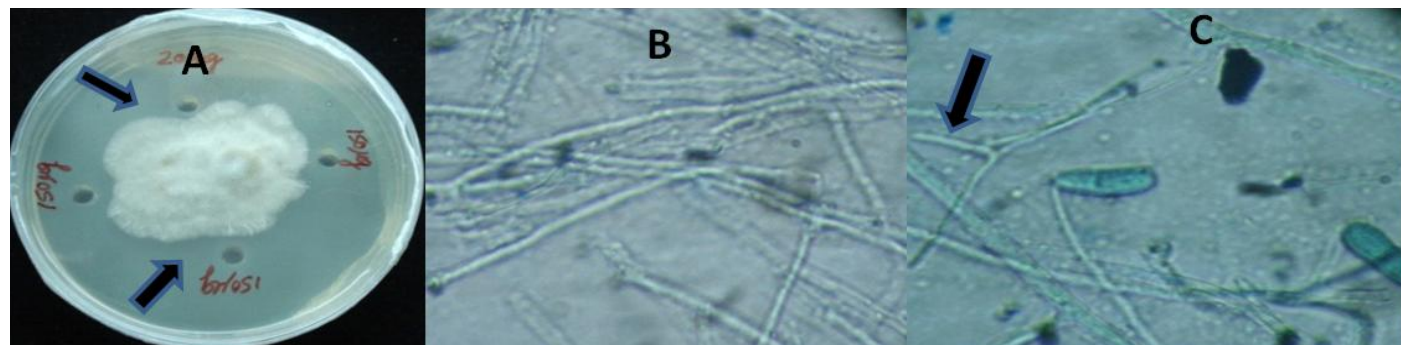

Figure 6. (A) Effect on growth of *Alternaria brassicicola* in response to different concentrations of endochitinase.Wells contain 50µl of enzyme solution in µgml^{-1} from top in clockwise direction (200,150,10050) (B+C) Enzyme disturbance seen in the form of branching but it is stable to recover as we did not find any sign of cellular bursting at 150- 200 µgml^{-1}.

stability of purified endochitinase. The enzyme activity was also determined at different pH (2 to 8). The highest activity was observed at pH 5 and enzyme was found stable from pH 5 to 7. These results are similar to the findings of Rashed et al. (2010) wherein the optimum temperature for *ech-42* at 35°C with maximum stability upto 55°C and highest activity at pH 6. Similarly there are other fungal chitinases (Chi 1,Chit 37,Chit 46) reported by
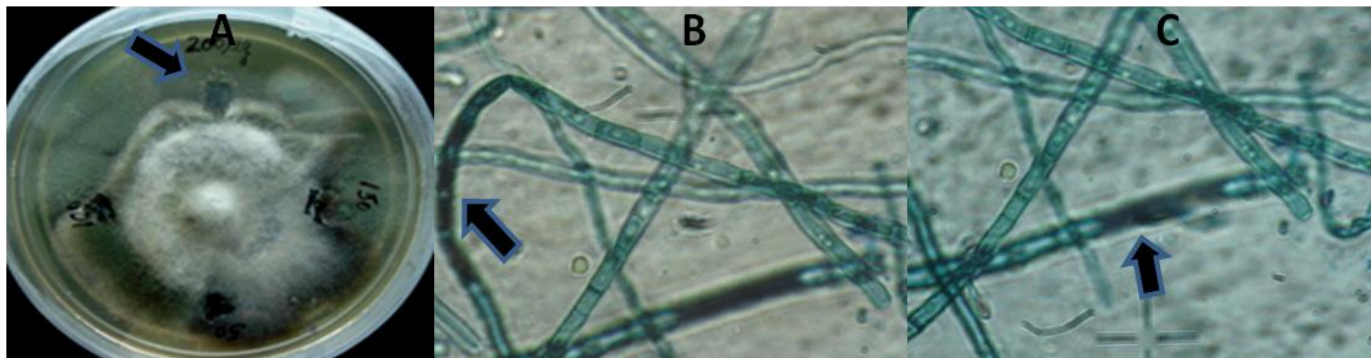
Figure 7. (A) Effect on growth of *Alternaria brassicae* in response to different concentrations of endochitinase. Wells contain 50 µl of enzyme solution in µgml^{-1} from top in clockwise direction (200, 150, 100 and 50). **(B)** Hyphal necrosis and swelling observed at concentration 200 µgml^{-1}. **(C)** Segmentation and enzyme disturbance observed.

several workers who found variable temperature and pH optimum (Gan et al., 2007; Caihong et al., 2007; Ike et al., 2006).

Investigations into the biological activity of *T. harzianum* clearly indicate strong reducing effect towards different phytopathogens *F. oxysporum* f.sp. *lycopersici*, *S. rolfsii*, *A. brassicae* and *A. brassicicola.* Microscopic examination revealed that the enzyme caused necrotic lesions, branching, hyphal bursting at the concentration of 200 µg ml^{-1} which is similar to the findings of Harjono and Widyastuti (2001), where pathogenic response of purified endochitinase of *T. reesei* on *Ganoderma philippii* was investigated. There are many workers who reported that the chitinases are substantially more active and effective against a wide range of fungi like Kaomek et al. (2003) who tested the antifungal activity of *L. leucocephala* chitinase against *Collectotrichum* sp., *Pestalestiopsis* sp. and *Fusarium* sp. and found it as good candidate for fungal inhibition.

Conclusion

It is concluded that 42-kDa endochitinase produced by *T. harzianum* has antifungal activity *in vitro*. The genes of *T. harzianum* coding for chitinolytic enzymes are attractive sources of these enzymes as their products can be used for combating with many phytopathogenic fungi. The enzymes like endochitinases can be used to add more biocontrol capabilities in wide series of microorganisms after further characterization. This could be an added advantage in bioprospecting novel antifungal or other compounds.

Conflict of Interests

The author(s) have not declared any conflict of interests.

ACKNOWLEDGEMENTS

REFERENCES

Alexandrine Froger, James E. Hall (2007). Transformation of Plasmid DNA into E. coli Using the Heat Shock Method. J. Vis. Exp. 6: 253. 2007.

Caihong H, Qian Y, Jinzhu S, Yingqi S (2007). Expression of a novel chitinase gene from *Trichoderma harzianum* in *Saccharomyces cerevisiae*. Proceedings of the 1st International Conference on Bioinformatics and Biomedical Engineering held at Wuhan, China. 1:283-285.

Dahiya N, Tewari R, Hoondal GS (2006). Biotechnological aspects of chitinolytic enzymes: a review. Appl. Microbiol. Biotechnol. 71: 773-782.

Gan Z, Yang J, Tao N, Liang L, Mi Q, Li J, Zhang Q (2007). Cloning of the gene *Lecanicillium psalliotae* chitinase *Lpchi1* and identification of its potential role in the biocontrol of root-knot nematode *Meloidogyne incognita*. Appl. Microbiol. Biotechnol. 76: 1309-1317.

Harighi MJ, Motallebi M, Zamani MR, (2007). Purification of chitinase 42 from *Trichoderma atroviride* PTCC5220. Iran. J. Biol. 19: 203-214.

Harjono, Widyastuti SM (2001).Antifungal activity of purified endochitinase prodced by biocontrol agent *Trichoderma reesei* against *Ganoderma philippii*. Pak. J. Biol. Sci. 4(10):1232-1234.

Harman GE (2006) Overview of mechanisms and uses of *Trichoderma* spp. Phytopathology 96: 190-194.

Howell CR (2003). Mechanisms employed by Trichoderma species in the biological control of plant diseases: The History and evolution of current concepts. Plant Dis. 87: 4-10.

Ike M, Nagamatsu K, Shioya A, Nogawa M, Ogasawara W, Okada H, Morikawa Y (2006). Purification, characterization, and gene cloning of 46 kDa chitinase (Chi46) from *Trichoderma reesei* PC-3-7 and its expression in *Escherichia coli*. Appl. Microbiol. Biotechnol. 71: 294-303.

Jhonson LF, Curl EA (1972). Methods for research on the ecology of soil-borne plant pathogens. Burgess Publishing company, Minnesota. pp. 241.

Kaomek M, Mizuno K,Fujimura T, Sriyotha P, Caima JRK (2003). Cloning expression and characterization of antifungal chitinase from *Leucaena leucocephala* de wilt. Biosci. Biotechnol. Biochem. 67:667-676.

Kim DJ, Baek JM, Uribe P, Kenerley CM, Cook DR (2002). Cloning and characterization of multiple glycosyl hydrolase genes from

Trichoderma virens. Curr. Genet. 40: 374-384.

Markovich NA, Kononova GL (2003). Lytic enzymes of Trichoderma and their role in plant defense from fungal diseases: a review. Appl. Biochem. Microb. 39(4): 341-351.

Merzendorfer H, Zimoch L (2003). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. J. Exp. Biol. 206: 4393-4412.

Miller GL (1959). Use of dinitrosalisylic acid reagent for determination of reducing sugars. Anal. Chem. 31:426-428.

Ningthoujam DS, Sanasam S, Nimaichand S (2009). Screening of actinomycete isolates from niche habitats in Manipur for antibiotic activity. Am. J. Biochem. Biotechnol. 5: 221-225.

Patil RS, Ghormade V, Deshpande MV (2000). Chitinolytic enzymes: an exploration. Enzyme Microb. Technol. 26: 473-483.

Rashed Al SAA, Bakar FDA, Said M, Hassan O Rabu A, Illias RM, Murad AMA (2010).Expression and characterization of the recombinant *Trichoderma virens* endochitinase Cht2. Afr. J. Microbiol. Res. 16:1758-1767.

Rattanakit N, Yano S, Plikomol A, Wakayama M, Tachiki T (2007). Purification of *Aspergillus* sp. S1-13 chitinases and their role in saccharification of chitin in mash of solid-state culture with shellfish waste. J. Biosci. Bioeng. 103: 535-541.

Salami SA, Ebadi A, Naghavi MR, Dry IB (2008). Cloning and functional characterization of a class III chitinase gene from grapevine: Inhibition of fungal growth by recombinant VvChiF III. Afr. J. Biotechnol. 7(24): 4400-4405.

Sharma P, Saravanan K, Ramesh R, Vignesh Kumar P, Dinesh Singh, Sharma M, Monica S. Henry and Swati Deep (2012). Cloning and Semi-quantitative gene expression of endochitinase (ech-42) gene from *Trichoderma spp*. Afr. J. Biotechnol. 11(66):12930-12938.

Sharma P, Vignesh Kumar P, Ramesh R, Saravanan K, Deep S, Sharma M, Mahesh S and Dinesh S (2011).Biocontrol genes from Trichoderma species : A Review. Afr. J. Biotechnol. 10(86):19898-19907.

Tunc TM, Kul M, Yaman H, Demirkaya E, Saldir M, Kurt I, Sarici SU, Alpay F (2008). Chitotriosidase activity in human milk from mothers of premature and full-term infants during the first month of lactation. Clin. Biochem. 41: 693-696.

Wen CM, Tseng CS, Cheng CY, Li YK (2005). Purification, characterization and cloning of a chitinase from Bacillus sp. NCTU 2. Biotechnol. Appl. Biochem. 35: 213-219.

African Journal of
Biotechnology

Related Journals Published by Academic Journals

- ■ *Biotechnology and Molecular Biology Reviews*
- ■ *African Journal of Microbiology Research*
- ■ *African Journal of Biochemistry Research*
- ■ *African Journal of Environmental Science and Technology*
- ■ *African Journal of Food Science*
- ■ *African Journal of Plant Science*
- ■ *Journal of Bioinformatics and Sequence Analysis*
- ■ *International Journal of Biodiversity and Conservation*