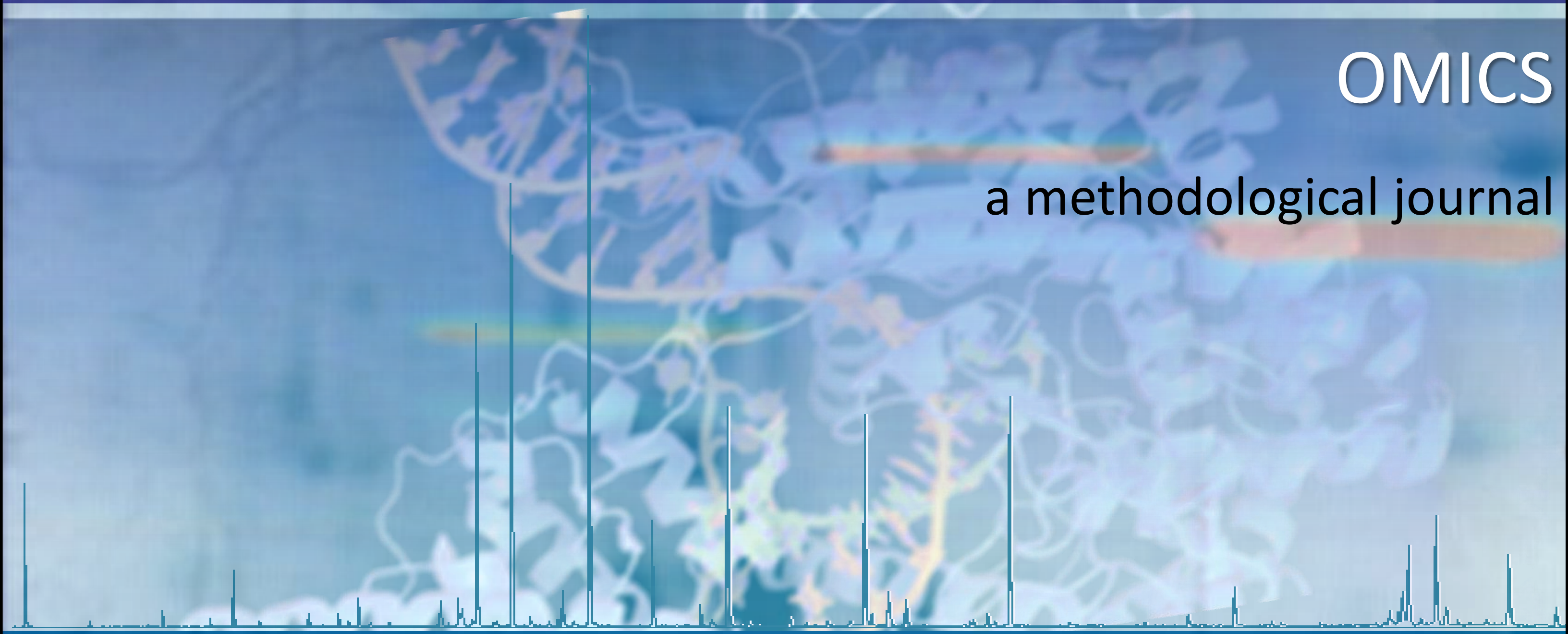
Volume 12 | Issue 1 | September 2022 ISSN 2182-0287

Journal of Integrated OMICS

a methodological journal



JIOMICS

Journal of Integrated OMICS

Focus and Scope

Journal of Integrated OMICS, JIOMICS, provides a forum for the publication of original research papers, preliminary communications, technical notes and critical reviews in all branches of pure and applied "-omics", such as genomics, proteomics, lipidomics, metabolomics or metallomics. The manuscripts must address methodological development. Contributions are evaluated based on established guidelines, including the fundamental nature of the study, scientific novelty, and substantial improvement or advantage over existing technology or method. Original research papers on fundamental studies, and novel sensor and instrumentation development, are especially encouraged. It is expected that improvements will also be demonstrated within the context of (or with regard to) a specific biological question; ability to promote the analysis of molecular mechanisms is of particular interest. Novel or improved applications in areas such as clinical, medicinal and biological chemistry, environmental analysis, pharmacology and materials science and engineering are welcome.

Editors-in-Chief

Carlos Lodeiro-Espiño, University NOVA of Lisbon, Portugal
Florentino Fdez-Riverola, University of Vigo, Spain
Jens R. Coorssen, Brock University, Ontario, Canada
Jose-Luís Capelo-Martínez, University NOVA of Lisbon, Portugal

Regional editors

ASIA

Nelson Cruz
Saudi Arabia

Europe

Gilberto Igrejas
University of Trás-os-Montes and Alto Douro, Life Sciences and Environmental School, Centre of Genetics and Biotechnology
Department of Genetics and Biotechnology, 5001-801 Vila Real, Portugal

Randen Patterson
Center for Computational Proteomics, The Pennsylvania State University, US

Jens R. Coorssen
Brock University, Ontario, Canada

South America

Marco Aurélio Zezzi Arruda
University of Campinas - Unicamp

Carlos H. I. Ramos
ChemistryInstitute – UNICAMP, Brazil

Associated editors

AFRICA

Saffaj Taouqif
Centre Universitaire Régional d'Interface, Université Sidi Mohamed Ben Abdallah, route d'Imouzzar-Fès, Morocco

ASIA

Amita Pal
Division of Plant Biology, Bose Institute, Kolkata, India
Ashish Gupta
Centre of Biomedical Magnetic Resonance, SGPGIMS Campus, India
Canhua Huang
The State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, PR China
Ching-Yu Lin
Institute of Environmental Health, College of Public Health, National Taiwan University, Taipei, Taiwan
Chantragan Srisomsap
Chulabhorn Research Institute, Bangkok, Thailand
Debmalya Barh
Institute of Integrative Omics and Applied Biotechnology (IIOAB), India

Eiji Kinoshita
Department of Functional Molecular Science, Graduate School of Biomedical Sciences, Hiroshima University, Japan
Fan Chen
Institute of Genetics and Developmental Biology, Chinese Academy of Sciences (CAS), China
Ganesh Chandra Sahoo
BioMedical Informatics Center of Rajendra Memorial Research Institute of Medical Science (RMRIMS), Patna, India
Guangchuang Yu
Institute of Life & Health Engineering, Jinan University, Guangzhou, China
Hai-Lei Zheng
School of Life Sciences, Xiamen University, China
Hsin-Yi Wu
Institute of Chemistry, Academia Sinica, Taiwan
Ibrokhim Abdurakhmonov
Institute of Genetics and Plant experimental Biology Academy of Sciences of Uzbekistan, Uzbekistan
Jianghao Sun
Food Composition and Method Development Lab, U.S. Dept. of Agriculture, Agricultural Research Services, Beltsville, USA
Juan Emilio Palomares-Rius
Forestry and Forest Products Research Institute, Tsukuba, Japan
Jung Min Kim
Liver and Immunology Research Center, Daejeon Oriental Hospital of Daejeon University, Republic of Korea
Kobra Pourabdollah
Razi Chemistry Research Center (RCRC), Shahreza Branch, Islamic Azad University, Shahreza, Iran
Krishnakumar Menon
Amrita Center for Nanosciences and Molecular Medicine, Amrita Institute of Medical Sciences, Kochi, Kerala, India
Mohammed Rahman
Center of Excellence for Advanced Materials Research (CEAMR), King Abdulaziz University, Jeddah, Saudi Arabia
Ningwei Zhao
Life Science & Clinical Medicine Dept. ; Shimadzu (China) Co., Ltd
Poh-Kuan Chong
National University of Singapore, Singapore
Sanjay Gupta
Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Kharghar, Navi Mumbai, India
Sanjeeva Srivastava
Indian Institute of Technology (IIT) Bombay, India
Suresh Kumar
Department of Applied Chemistry, S. V. National Institute of Technology, Gujarat, India
Toshihide Nishimura
Department of Surgery I, Tokyo Medical University, Tokyo, Japan
Vishvanath Tiwari
Department of Biochemistry, Central University of Rajasthan, India
Xuanxian Peng
School of Life Sciences, Sun Yat-sen University, Guangzhou, China
Youxiong Que
National Research & Development Center for Sugarcane, China Agriculture Research System(CARS), Fujian Agriculture & Forestry University, Republic of China
Yu Wang
Department of Pharmacology and Pharmacy, the University of Hong Kong, China
Zhiqiang Gao
Department of Chemistry, National University of Singapore

AUSTRALIA AND NEW ZEALAND

Emad Kiriakous
Queensland University of Technology (QUT), Brisbane, Australia
Joëlle Coumans-Moens
School of Science and Technology, School of Medicine, University of New England, Australia
Maurizio Ronci
Mawson Institute, University of South Australia, Mawson Lakes, Australia
Michelle Colgrave
CSIRO Livestock Industries, St Lucia, Australia
Peter Hoffmann
Institute for Photonics & Advanced Sensing (IPAS), School of Chemistry and Physics, University of Adelaide, Australia
Valerie Wasinger
Bioanalytical Mass Spectrometry Facility, Mark Wainwright Analytical Centre, University of NSW, Australia
Wujun Ma
Centre for Comparative Genomics, Murdoch University, Australia

EUROPE

AhmetKoc, PhD
Izmir Institute of Technology, Department of Molecular Biology & Genetics, Urla, İzmir, Turkey
Alejandro Gella
Department of Basic Sciences, Neuroscience Laboratory, Faculty of Medicine and Health Sciences, Universitat Internacional de Catalunya, Sant Cugat del Vallès-08195, Barcelona, Spain
Angelo D'Alessandro
Università degli Studi della Tuscia, Department of Ecological and Biological Sciences, Viterbo, Italy
Antonio Gnoni
Department of Medical Basic Sciences, University of Bari "Aldo Moro", Bari, Italy
Ana Varela Coelho
Instituto de Tecnologia Química e Biológica (ITQB) Universidade Nova de Lisboa (UNL), Portugal
Anna Maria Timperio
Dipartimento Scienze Ambientali Università della Tuscia Viterbo, Italy
Andrea Scaloni
Proteomics and Mass Spectrometry Laboratory, ISPAAM, National Research Council, via Argine 1085, 80147 Napoli, Italy
Angel P. Diz
Department of Biochemistry, Genetics and Immunology, Faculty of Biology, University of Vigo, Spain
Angela Chambery
Department of Life Science, Second University of Naples, Italy
Anna-Irini Koukkou
University of Ioannina, Department of Chemistry, Biochemistry Laboratory, Greece
António Sebastião Rodrigues
Departamento de Genética, Faculdade de Ciências Médicas, Universidade Nova de Lisboa,Portugal
Arzu Umar
Department of Medical Oncology, Laboratory of Breast Cancer Genomics and Proteomics, Erasmus Medical Center Rotterdam Josephine Nefkens Institute, Rotterdam, The Netherlands
Bart Devreese
Laborartory for Protein Biochemistry and Biomolecular Engineering, Department for Biochemistry and Microbiology, Ghent University, Belgium
Bernard Corfe
Department of Oncology, University of Sheffield, Royal Hallamshire Hospital, United Kingdom
Bruno Manadas
Center for Neuroscience and Cell Biology, University of Coimbra, Portugal
Carla Pinheiro
Plant Sciences Division, Instituto de Tecnologia Química e Biológica (ITQB), Universidade Nova de Lisboa, Portugal
Claudia Desiderio
Consiglio Nazionale delle Ricerche, Istituto di Chimica del Riconoscimento Molecolare (UOS Roma), Italy
Claudio De Pasquale
SAgA Department, University of Palermo, Italy
Celso Vladimiro Cunha
Medical Microbiology Department, Institute of Hygiene and Tropical Medicine, New University of Lisbon, Portugal

Christian Lindermayr
Institute of Biochemical Plant Pathology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
Christiane Fæste
Section for Chemistry and Toxicology Norwegian Veterinary Institute, Oslo, Norway
Christophe Cordella
UMR1145 INRA, Laboratoire de Chimie Analytique, Paris, France
Cosima Damiana Calvano
Universita' degli Studi di Bari, Dipartimento di Chimica, Bari, Italy
Daniela Cecconi
Dip. diBiotecnologie, LaboratoriodiProteomica e Spettrometriadi Massa, Universitàdi Verona, Verona, Italy
Deborah Penque
Departamento de Genética, Instituto Nacional de Saúde Dr Ricardo Jorge (INSA, I.P.), Lisboa, Portugal
Dilek Battal
Mersin University, Faculty of Pharmacy, Department of Toxicology, Turkey
Domenico Garozzo
CNR ICTP, Catania, Italy
Ed Dudley
Institute of Mass Spectrometry, College of Medicine Swansea University, Singleton Park, Swansea, Wales, UK
Elia Ranzato
Dipartimento di Scienze e Innovazione Tecnologica, DiSIT, University of Piemonte Orientale, Alessandria, Italy
Elisa Bona
Università del Piemonte Oientale, DISIT, Alessandria, Italy
Elke Hammer
Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt Universität, Germany
Enrica Pessione
University of Torino, Life Sciences and Systems Biology Department, Torino, Italy
Federica Pellati
Department of Life Sciences, University of Modena and Reggio Emilia, Italy
François Fenaille
CEA, IBiTecS, Service de Pharmacologie et DImmunoanalyse (SPI), France
Fulvio Magni
Department of Health Science, Monza, Italy
Georgios Theodoridis
Department of Chemistry, Aristotle University, Greece
Gianfranco Romanazzi
Department of Environmental and Crop Sciences, Marche Polytechnic University, Italy
Giorgio Valentini
Università degli Studi di Milano, Dept. of Computer Science, Italy
Helen Gika
Chemical Engineering Department, Aristotle University of Thessaloniki, Greece
Hugo Miguel Baptista Carreira dos Santos
REQUIMTE-FCT Universidade NOVA de Lisboa, Portugal
Iñaki Álvarez
Institut de Biotecnologia i Biomedicina Vicent Villar Palasí, Universitat Autònoma de Barcelona, Barcelona
Jane Thomas-Oates
Centre of Excellence in Mass Spectrometry and Department of Chemistry, University of York, Heslington, UK
Jens Allmer
Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Izmir, Turkey
Jesús Jorrín Novo
Agricultural and Plant Biochemistry, Proteomics Research Group, Department of Biochemistry and Molecular Biology, Córdoba, Spain
Johan Palmfeldt
Research Unit for Molecular Medicine, Aarhus University Hospital, Skejby, Aarhus, Denmark
Jose Câmara
University of Madeira, Funchal, Portugal
Juraj Gregan
Max F. Perutz Laboratories, University of Vienna, Austria
Karin Stensjö
Department of Photochemistry and Molecular Science, Ångström laboratory, Uppsala University, Sweden
Kay Ohlendieck
Department of Biology, National University of Ireland, Maynooth, Co. Kildare, Ireland
Konstantinos Kouremenos
Department of Chemistry, Umea University, Sweden
Luisa Brito
Laboratório de Microbiologia, Instituto Superior de Agronomia, Tapada da Ajuda, Lisbon, Portugal
Marco Lemos
GIRM & ESTM - Polytechnic Institute of Leiria, Peniche, Portugal
María Álava
Departamento de Bioquimica y Biologia Molecular y Celular, Facultad de Ciencias, Universidad de Zaragoza, Spain
María de la Fuente
Legume group, Genetic Resources, Mision Biologica de Galicia-CSIC, Pontevedra, Spain
Maria Gabriela Rivas
REQUIMTE/CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
Marie Arul
Muséum National Histoire Naturelle, Département RDDM, Plateforme de spectrométrie de masse et de protéomique, Paris, France
Marie-Pierre Bousquet
Institut de Pharmacologieet de Biologie Structurale, UPS/CNRS, Tolouse, France
Mario Diniz
Dept. Química-REQUIMTE, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
Martina Marchetti-Deschmann
Institute of Chemical Technologies and Analytics, Vienna University of Technology, Vienna, Austria
Maxence Wisztorski
University Lille 1, Laboratoire de Spectrométrie de Masse Biologique, Fondamentale & Appliquée, Villeneuve d'ascq, France
Michel Jaquinod
Exploring the Dynamics of Proteomes/Laboratoire Biologie à Grande Echelle, Institut de Recherches en Technologies et Sciences pour le Vivant, Grenoble, France
Mónica Botelho
Centre for the study of animal sciences (CECA)/ICETA, Porto, Portugal
Pantelis Bagos
Department of Computer Science and Biomedical Informatics, University of Central Greece, Greece
Patrice Francois
Genomic Research Laboratory, Service of Infectious Diseases, Department of Internal Medicine, Geneva
Patrícia Alexandra Curado Quintas Dinis Poeta
University of Trás-os-Montes and Alto Douro (UTAD), School of Agrary and Veterinary Sciences, Veterinary, Science Department, Portugal
Pedro Rodrigues
Centro de Ciências do Mar do Algarve, CCMAR, Faro, Portugal
Per Bruheim
Department of Biotechnology, Norwegian University of Science and Technology, Trondheim, Norway
Philipp Hess
Institut Universitaire Mer et Littoral(CNRS - Université de Nantes - Ifremer), Nantes, France
Pieter de Lange
DipartimentodiScienzedellaVita, SecondaUniversità degli Studi di Napoli, Caserta, Italy
Ralph Fingerhut
University Children`s Hospital, Swiss Newborn Screening Laboratory, Children`s Research Center, Zürich, Switzerland
Rubén Armañanzas
Computational Intelligence Group, Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Spain
Ruth Birner-Gruenberger
Medical University Graz, Austria

Sebastian Galuska
Institute of Biochemistry, Faculty of Medicine, Justus-Liebig-University of Giessen, Germany
Serge Cosnier
Department of Molecular Chemistry, Grenoble university/CNRS, Grenoble, France
Serhat Döker
Cankiri Karatekin University, Chemistry Department, Cankiri, Turkey
Simona Martinotti
Dipartimento di Scienze e Innovazione Tecnologica, DiSIT, University of Piemonte Orientale, Alessandria, Italy
Spiros D. Garbis
Biomedical Research Foundation of the Academy of Athens, Center for Basic Research - Division of Biotechnology, Greece
Steeve Thany
Laboratoire Récepteurs et Canaux Ioniques Membranaires, UFR Science, Université d'Angers, France
Stefania Orrù
University if Naples Parthenope, Naples, Italy
Tâmara García Barrera
Departamento de Química y Ciencia de losMateriales, Facultad de Ciencias Experimentales, Universidad de Huelva, Spain
Vera Muccilli
DipartimentodiScienzeChimiche, UniversitàdiCatania, Catania, Italy
Yuri van der Burgt
Leiden University Medical Center, Department of Parasitology, The Netherlands

SOUTH AMERICA

Andréa P.B. Gollucke
Hexalab/Catholic University of Santos, Brazil
Arlindo Moura
Department of Animal Science - College of Agricultural Sciences - Federal University of Ceara, Fortaleza, Brasil
Bruno Lomonte
Instituto Clodomiro Picado, Universidad de Costa Rica
Edson Guimarães Lo Turco
São Paulo Federal University, Brasil
Fabio Ribeiro Cerqueira
Department of Informatics and NuBio (Research Group for Bioinformatics), University of Vicosa, Brazil
Fernando Barbosa
Faculty of Pharmaceutical Sciences of Ribeirão Preto University of São Paulo, Brazil
Mário Hiroyuki Hirata
Laboratório de Biologia Molecular Aplicado ao Diagnóstico, Departamento de Análises Clínicas e Toxicológicas, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, Brazil
Jan Schripsema
Grupo Metabolômica, Laboratório de Ciências Quimicas, Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, Brazil
Jorg Kobarg
Centro Nacional de Pesquisa em Energia e Materiais, Laboratório Nacional de Biociências, Brazil
Rossana Arroyo
Department of Infectomic and Molecular Biology, Center of Research and Advanced Studies of the National, Polytechnical Institute (CINVESTAV-IPN), Mexico City, Mexico
Rubem Menna Barreto
Laboratorio de Biología Celular, Instituto Oswaldo Cruz, Fundaçao Oswaldo Cruz, Rio de Janeiro, Brazil
Vasco Azevedo
BiologicalSciencesInstitute, Federal University of Minas Gerais, Brazil

NORTH AMERICA

Amosy M'Koma
Vanderbilt University School of Medicine, Department of General Surgery, Colon and Rectal Surgery, Nashville, USA
Anthony Gramolini
Department of Physiology, Faculty of Medicine, University of Toronto, Canada
Anas Abdel Rahman
Department of Chemistry, Memorial University of Newfoundland and Labrador St. John's, Canada
Christina Ferreira
Purdue University - Aston Laboratories of Mass Spectrometry, Hall for Discovery and Learning Research, West Lafayette, US
Eustache Paramithiotis
Caprion Proteomics Inc., Montreal, Canada
Jagjit Yadav
Microbial Pathogenesis and Toxicogenomics, Laboratory, Environmental Genetics and Molecular, Toxicology Division, Department of Environmental
Health, University of Cincinnati College of Medicine, Ohio, USA
Jiaxu Li
Department of Biochemistry and Molecular Biology, Mississippi State University, USA
Laszlo Prokai
Department of Molecular Biology & Immunology, University of North Texas Health Science Center, Fort Worth, USA
Madhulika Gupta
Children's Health Research Institute, University of Western Ontario London, ON, Canada
Michael H.A. Roehrl
Department of Pathology and Laboratory Medicine, Boston Medical Center Boston, USA
Olgica Trenchevska
Molecular Biomarkers, Biodesign Institute at Arizona State University, USA
Robert Powers
University of Nebraska-Lincoln, Department of Chemistry, USA
Susan Hester
United Stated Environmental Protection Agency, Durnam, USA
Thomas Kislinger
Department of Medical Biophysics, University of Toronto, Canada
William A LaFramboise
Department of Pathology, University of Pittsburgh School of Medicine Shadyside Hospital, Pittsburgh, USA
Xuequn Chen
Department of Molecular & Integrative Physiology, University of Michigan, Ann Arbor, USA
Ying Qu
Microdialysis Experts Consultant Service, San Diego, USA

JOURNAL OF INTEGRATED OMICS

A methodological Journal

EDITORIAL

ORIGINAL ARTICLES

EDITORIAL LETTER

JIOMICS | VOL 12 | ISSUE 1 | June 2022 | 211 | 1-17

JOURNAL OF INTEGRATED OMICS

A METHODOLOGICAL JOURNAL

HTTP://WWW.JIOMICS.COM

EDITORIAL | DOI: 10.5584/jiomics.v12i1.211

Prescriptomics: the next frontier in medicine

José L. Capelo[1, 2]*, Carlos Lodeiro[1,2], Rajiv Dhir[3], Nataly Kravchenko-Balasha[4], Hugo M. Santos[1,2,3]*

[1] (Bio)Chemistry & Omics, BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516, Caparica, Portugal; [2] PROTEOMASS Scientific Society, Madan Parque, Rua dos Inventores, 2825-182 Caparica, Portugal; [3] Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA, United States; [4] Department for Bio-medical Research, Institute of Dental Sciences, Hebrew University of Jerusalem, Jerusalem, 91120, Israel

Available Online: 23 May 2022

Analysis of proteins has been an integral part of clinical chemistry for decades. However, recent technological advances have opened new opportunities for the large-scale analysis of proteins for clinical diagnostic purposes and personalized medicine [1]. First, the development of mass spectrometers with significantly higher resolution and larger dynamic range has allowed generating high-quality quantitative data from complex sample matrices such as serum, plasma, urine, cerebrospinal fluid, and tissue biopsies without the need for isotopic labellings [2, 3]. Instruments like TOF, Orbitrap, FT -ICR and the most recent timsTOF [4, 5] were a stepping-stone towards proteomic-based personalized medicine by facilitating the detection of patient-specific protein signatures that reorganize over time due to genetic, environmental, and treatment constraints.

Second, the development of high-throughput robust sample preparation methods has allowed the processing of many clinical specimens in an efficient and reproducible manner [6]. Parallel to such achievements, the exciting advances experienced by bioinformatics have been a cornerstone in translating quantitative proteomics data into actionable clinical information [7]. Thus, the development of new and powerful algorithms, artificial intelligence, and information theoretics, combined with large datasets available in data repositories such as TCGA [8] or Proteome exchange [9] are moving biology and medicine where no other sciences have gone before [10]. Nowadays, the proteomics community is addressing a critical question: is it possible to obtain helpful information from these datasets to determine which combination of FDA or EMA-approved drugs is best suited to treat each patient? [11]. Exciting discoveries are expected from re-analyzing these already available datasets.

The concept of genomics soon evolved into the concept of phenome [12], the latter referring to the actual expression of an individual's genes. As phenomics evolves and matures, it is becoming clear that pre-defined categories where diseases were classified and grouped, contradict the vast heterogeneity of phenotypes. Thus, the treatment of phenotype-dependent complex diseases with extensive heterogeneity, like cancer, is likely to overlook unique patients that must be treated individually. We are what we express. And what we express can be now quantified with unprecedented precision and accuracy. By quantifying the proteome expressed by a patient, any disease can be diagnosed, prognosed and followed either using a tissue or in a liquid biopsy (urine, serum, plasma). However, a challenge remains. Can the information provided by proteomics and phenomics be translated to prescriptomics? This is, based on patient-specific alterations in protein-protein networks will prescription be personalized? Yes, proteomics will be a good bridge between diagnostics and therapeutics. This is our best answer based on our current knowledge on disease, genomics, proteomics, and phenomics [11–16].

So far, this is how we visualized the revolution in medicine ahead of our time. Within the next five to ten years, the ultrafast analysis of proteomes of liquid and solid biopsies via high-resolution mass spectrometry will be done in minutes. Currently, mass spectrometry technologies allow

***Corresponding author:** José L. Capelo, email: jlcm@fct.unl.pt; Hugo M. Santos, email: hmsantos@fct.unl.pt

for the absolute quantification of hundreds of proteins in just 15 min. [16, 17]. Next, the bioinformatic analysis of the sampled proteome(s) will deliver information about the altered biological pathways in each patient, for instance, to uncover whether specific signalling pathways or the inflammatory and immune response pathways are dysregulated [18, 19]. Also, it will collect information about single and well-characterized biomarkers to complement the information taken from the interpretation of the proteome.

Altogether, each unique phenotype will lead to a personalized prescription based on a diagnosis done at a protein-pathway interaction level. This is what we call prescriptomics. Further, linking this technology with communication via electronic interface and drone delivery will make physician and patient interaction faster. Within the next ten years, technological tools will be implemented, allowing proteomic analysis to be done at home just using saliva or urine, not to mention blood. These tools are currently used in hospitals, but interfaces making them ready to use at home are soon expected. By accomplishing the tool to the computer or the mobile phone, the information will be delivered to the physician via a software interface addressing a potential prescription. In case of need, specific tools or sample collecting devices will be provided to the patient and picked up by drones, then transporting again the sample to the laboratory of analysis. Such an approach brings several advantages. For instance, a lower number of patients attending the hospital face to face, thus avoiding the risk of becoming infected or vectors spreading diseases. Also, a most effective medical decision making is expected. Furthermore, the following up of the patient's proteome during and after medical care will thus be made easier and will allow the implementation of dynamic therapy as the standard method to treat disease providing a unique system to alert for relapses.

Acknowledgment

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura (#PM001/2019 and #PM003/2016). The authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV, which is financed by national funds from FCT/MCTES (UIDB/50006/2020) and UIDP/50006/2020). H. M. S. acknowledges FCT/MCTES (Portugal) for his research contract FCT Investigator Program 2015 (IF/00007/2015).

References

1] K.K. Jain. Advances in Protein Chemistry and Structural Biology (2016), 102, 41–52. DOI: 10.1016/BS.APCSB.2015.09.002

[2] T. Sajic, Y. Liu, R. Aebersold. PROTEOMICS – Clinical Applications (2015), 9, 307–321. DOI: 10.1002/PRCA.201400117

[3] S. Jorge, J.L. Capelo, W. LaFramboise, S. Satturwar, D. Korentzelos, S. Bastacky, G. Quiroga-Garza, R. Dhir, J.R. Wiśniewski, C. Lodeiro, H.M. Santos. BMC Medicine (2021), 19, 1–13. DOI: 10.1186/S12916-021-02071-9

[4] M. Hardman, A.A. Makarov. Analytical Chemistry (2003), 75,1699–1705. DOI: 10.1021/AC0258047

[5] F. Meier, A.D. Brunner, S. Koch, H. Koch, M. Lubeck, M. Krause, N. Goedecke, J. Decker, T. Kosinski, M.A. Park, N. Bache, O. Hoerning, J. Cox, O. Räther, M. Mann. Molecular and Cellular Proteomics 2018, 17, 2534–2545. DOI: 10.1074/MCP.TIR118.000900

[6] L.B. Carvalho, J.L. Capelo-Martínez, C. Lodeiro, J.R. Wiśniewski, H.M. Santos. Analytical Chemistry (2020), 92, 13, 9164–9171. DOI: 10.1021/acs.analchem.0c01470

[7] M. Manfredi, J. Brandi, C. di Carlo, V. Vita Vanella, E. Barberis, E. Marengo, M. Patrone, D. Cecconi. Expert Review of Proteomics, (2019), 16:9, 733-747. DOI: 10.1080/14789450.2019.1654862

[8] https://www.cancer.gov/tcga. Last time accessed 23rd May 2022

[9] http://proteomecentral.proteomexchange.org/cgi/GetDataset. Last time accessed 23rd May 2022

[10] J. He, S.L. Baxter, J. Xu, J. Xu, X. Zhou, K. Zhang. Nature Medicine 2019, 25:1, 30–36. DOI: 10.1038/s41591-018-0307-0

[11] S. Vasudevan, E. Flashner-Abramson, Heba Alkhatib, Sangita Roy Chowdhury, I. A. Adejumobi, D. Vilenski, S. Stefansky, A. M. Rubinstein & N. Kravchenko-Balasha. Precision Oncology (2021), 5, 50. DOI 10.1038/s41698-021-00190-3

[12] M.A. Bogue, S.C. Grubb. Genetica (2004), 122:1, 71–74. DOI: 10.1007/S10709-004-1438-4

[13] E. Flashner-Abramson, S. Vasudevan, I.A. Adejumobi, A. Sonnenblick, N. Kravchenko-Balasha. Theranostics (2019), 9, 5149–5165. DOI: 10.7150/THNO.31657

[14] S. Vasudevan, I.A. Adejumobi, H. Alkhatib, S.R. Chowdhury, S. Stefansky, A.M. Rubinstein, N. Kravchenko-balasha, Drug-Induced Resistance and Phenotypic Switch in Triple-Negative Breast Cancer Can Be Controlled via Resolution and Targeting of Individualized Signaling Signatures. Cancers 2021, 5009, 13. DOI: 10.3390/CANCERS13195009

[15] N. Kravchenko-Balasha. PROTEOMICS 2020, 20, 1900227. DOI: 10.1002/PMIC.201900227

[16] https://youtu.be/J0uGhyVDiuQ. Last time accessed 23rd May 2022

[17] https://www.bruker.com/content/bruker/int/en/products-and-solutions/mass-spectrometry/timstof/timstof-pro.html. Last time accessed 23rd May 2022

[18] L.B. Carvalho, J.L. Capelo, C. Lodeiro, R. Dhir, L.C. Pinheiro, M. Medeiros, H.M. Santos, BioRxiv (2021) 2021.04.30.442131. DOI: 10.1101/2021.04.30.442131

[19] https://www.youtube.com/watch?v=8DAZq40wmMo. Last time accessed 23rd May 2022.

ORIGINAL ARTICLES

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ARTICLE | DOI: 10.5584/jiomics.v12i1.212

Comparative analyses of the complete mitochondrial genomes of the two filarial worms *Wuchereria bancrofti & Brugia malayi* with *Caenorhabditis elegans*

Saranya Joshi[1], Lokesh Kumar[1], Kanchan Rauthan[1], Sudhir Kumar[1]*

[1] Department of Biotechnology, Hemvati Nandan Bahuguna Garhwal University (A Central University), Srinagar Garhwal, Uttarakhand, India.

Received: 14 July 2022 **Accepted:** 13 September 2022 **Available Online**: 30 September 2022

ABSTRACT

Wuchereria bancrofti and *Brugia malayi* are filarial worms belonging to the phylum Nematoda and cause lymphatic filariasis (LF) disease in humans. *W. bancrofti* and *B. malayi* are *Wolbachia* dependent organisms while *C. elegans* is a free living *Wolbachia* independent nematode. To investigate the conserved regions present in the mitochondrial genome of these organisms, the complete mitochondrial (mt) genomes of *W. bancrofti* and *B. malayi* having sizes 13,636 bp and 13,657 bp in length, respectively are compared with *C. elegans* (13794 bp). These mt genomes were similar to each other in respect of their size, and AT content and encode the same 12 PCGs (nad1–6, nad4L, cytb, cox1–3, and atp6). Complete mt genome alignment identified 13 conserved regions in each of the organisms with some of these regions unique only to one organism. Phylogenetic analysis using the mt genome showed a close relationship between *W. bancrofti* and *B. malayi* but showed a common early ancestor with the *C. elegans* emphasizing an early evolutionary divergence.

Keywords: *Comparative genomics; Mitochondrial DNA; conserved regions; Phylogenetic analysis; ORFs .*

1. Introduction

Wuchereria bancrofti and *Brugia malayi* are parasitic filarial worms belonging to the phylum Nematoda and cause lymphatic filariasis (LF) in humans. Worldwide, LF is a neglected tropical disease [1] affecting thousands of individuals and has little research investment [2]. This disease is not lethal but its infection is chronic leading to lifetime deformity which is not curable and its treatment is also difficult [3]. The symptoms including lymphoedema, elephantiasis and scrotal swelling are painful and can lead to permanent physical impairment. Along with impairment, the patient suffers from economical losses and the psychological stigma related to it is also devastating in the sense of social as well as self-acceptance (WHO factsheet). The current focus is on controlling its transmission, but an in-depth exploration of its genomic sequences can provide the baseline molecular data on these parasites giving insight into their evolutionary path.

The genomic study of an individual is done either with its nuclear or mitochondrial (mt) genome or with both. This study deals with a comparative analysis of mitochondrial genomes which contains a minor but important component of a eukaryote's genome [4]. Even though animal mt sequences are known to evolve rapidly, their gene arrangements typically remained constant throughout their evolution [5]. That is why the mt genome has always been extensively used as a molecular marker for phylogenetic studies [1]. In all animals, the mt genomes, with a few exceptions, include the same 37 genes: 13 encoding for proteins, 2 for rRNAs, and 22 for tRNAs. The variation that exists in mitochondrial genome size is usually related to the length differences of non-coding regions, the repetitive nature of some sequences, and/or the presence of huge duplications in some species [6]. Parasitic nematodes

---

***Corresponding author:** Dr. Sudhir Kumar—Department of Biotechnology, Chauras Campus, H.N.B. Garhwal University, Srinagar (Garhwal)-246174, Uttarakhand, India. Email: sudhir.1685@gmail.com

have a compact, circular mt genome which varies in size from 13,000 to 26,000 bp [7]. The selected nematodes under this study have mt genome with 12 protein-coding genes (PCGs) namely nad1–6, cox1–3, nad4L, cytb, and atp6 (lacking atp8 gene), 2 rRNA genes, and 22 tRNA genes [6]. The comparative study of all these genes will help search for the resemblances and variations between them at the genomic level.

Comparative genomics is not only a tool but also a comprehensive approach in bioinformatics research for comparing species to find out similarities and differences at the sequence level. Its principle involves genes and genomes to be studied and compared in the phylogenetic context of the evolutionary process[8]. In this process, we exploit the fact that the genes encoding similar characters are part of conserved DNA between two species. The conserved DNA sequences, which encode functional proteins must remain preserved from the last common ancestor up to the current generation [9].

The purpose of this study is to perform comparative genomics of the mt genome of parasitic nematodes *Wuchereria bancrofti* and *Brugia malayi* with free-living nematode *Caenorhabditis elegans*. Since both the *W. bancrofti* and *B. malayi* are *Wolbachia*-dependent parasites [10], they are supposed to have some variations as compared to the free-living nematode *C. elegans* at the genomic level. Researchers previously compared the mt genomes of *Wolbachia* dependent and independent nematodes but could not find any major difference [11]. We hypothesized that the conserved regions in the mitochondrial genomes of these organisms may reveal the crucial differences or similarities that may help in understanding their nature, their genomic features, and their phylogenetic relationships with each other.

2. Material and Methods

2.1 Data retrieval

The mt genomes of the species *Wuchereria bancrofti* (GenBank accession number NC_016186), *Brugia malayi* (GenBank accession number AF538716.1), and *Caenorhabditis elegans* (GenBank accession number NC_001328) were retrieved from NCBI. The mt nucleotide content (ATGC) of all the three species was calculated using GC Content Calculator, online analysis, and plot tool provided by BiologicsCorp (https://www.biologicscorp.com/tools/GCContent/index).

2.2 Identification of conserved regions

The complete mt genomic sequences of the species *W. bancrofti*, *B. malayi* and *C. elegans* were aligned using an open-source MAUVE aligner, version 2.3.1, via progressive algorithm [12]. In Mauve software, the conserved sequences of genomes are aligned with rearrangements. First, the mt genomic sequence of *W. bancrofti* was aligned with *C. elegans*. It was followed by alignment of the mt genomic sequence of *B. malayi* with *C. elegans* and *W. bancrofti* respectively.

The sequences identified as conserved regions by mauve software were further analyzed against the NCBI database using BLASTP. In each conserved region identified, Open Reading Frames (ORFs) were predicted by using the ORFfinder tool of NCBI (https://www.ncbi.nlm.nih.gov/orffinder/) followed by smart BLAST analysis for functional identification.

2.3 Identification of conserved intergenic regions

The identification of conserved intergenic regions were performed by the alignment of complete mt genome sequence of *W. bancrofti* and *B. malayi* with *C. elegans* individually and at last *W. bancrofti* and *B. malayi* were aligned with each other. The species were aligned with the wgVISTA tool of the Vista server [13]. Similar alignment of the mitochondrial genomes of all these three species was also done by UCSC genome browser [14].

2.4 ORF prediction

The ORFs were predicted in the complete mt genome sequences of all the species using the ORFfinder tool of NCBI (https://www.ncbi.nlm.nih.gov/orffinder/). Each ORF predicted by the ORFfinder tool of NCBI was further subjected to BLASTP (protein-protein BLAST) to find out its probable function.

2.5 Multiple sequence alignment

Multiple Sequence alignment (MSA) of all three species was performed using the CLUSTALW server of EMBL-EBI [15]. For this, the slow and accurate pairwise alignment parameter was adopted.

2.6 Phylogenetic Analysis

Phylogenetic analysis of these species was done along with 59 other nematode species by using their complete mt genomic sequences in Mega X [16]. Neighbor-Join and BioNJ algorithms were applied to obtain the Initial tree(s) by using the Maximum Composite Likelihood (MCL) approach. The analysis involved 62 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There was a total of 26194 positions in the final dataset.

2.7 Comparison of protein coding genes

All the 12 protein coding genes (PCGs) namely nad1–6, nad4L, cytb, cox1–3, and atp6 found in the mt genome of the three species, along with their amino acid sequence of proteins were retrieved from the NCBI database and aligned with each other using CLUSTALW [15] and their alignment score were recorded to determine the sequence identity in them. The amino acid sequence of proteins were cross referred with their sequence in PDB database. The amino acid sequence reterived from NCBI and PDB database were aligned by pBLAST to validate the sequence similarity.

3. Results

3.1 Genomic Details

The complete mt genomes of *W. bancrofti* (GenBank accession number: AP017705.1), *B. malayi* (GenBank accession number AF538716.1), and *C. elegans* (GenBank accession number NC_001328.1) are having size 13636 bp, 13657 bp, and 13794 bp respectively. *W. bancrofti* has the smallest genome among the three species. The nucleotide content of all three mt genomes is enlisted (Table 1). *W. bancrofti* and *B. malayi* has almost similar AT content while *C. elegans* has a slightly more percentage of A nucleotide.

3.2 Identification of conserved regions

With help of results generated by MAUVE software (Figure 1), the output backbone file was used for the quantitative identification of conserved genomic order. The information regarding the coordinates of the regions shared by two genomes and their localization in the genomes is present within it. The ORFs were located in each of the conserved regions.

In the case of alignment of mt genomes of *W. bancrofti* to *C. elegans*, a total of 13 conserved regions have been identified separately for each genome. Out of which, 7 regions were aligned to each other while the remaining 6 regions were not (suggesting it to be unique between them). Among all the conserved regions, a maximum of 12 ORFs were predicted in the 4137 to 7039 nucleotide region of *C. elegans*. It also identified 5 regions where no ORF was predicted. In the same alignment result, 20 ORFs were found in the 3095-6932 nucleotide region of *W. bancrofti* while 2 regions were identified to have no ORFs. Most of the ORFs showed functional similarity with the 12 protein coding genes of their respective genome. Two hypothetical proteins namely WUBG_19143 (of *W. bancrofti* in the 1574-1819 nucleotide region) and DI535_26975 (of *C. elegans* in the 8655-9429 nucleotide region) have also been identified. In the 12153-13187 nucleotide region of *W. bancrofti*, one ORF showed functional similarity with the neuronal IgCAM of *C. elegans*. A multiple C2 and transmembrane domain-containing protein 2 isoform X1 has been predicted in one of the ORF of 7029-10006 nucleotide region of *W. bancrofti* (Table 2).

In the alignment result of *B. malayi* to *C. elegans*, 13 conserved regions were identified in which 7 regions were aligned to one another while the remaining 6 were unique between them. A maximum of 25 ORFs were predicted in the 2590-9950 nucleotide region of *C. elegans* while 7 regions didn't show any ORF. Similarly, 6 regions of *B. malayi* didn't have any ORF in them. No full-length PCGs in ORFs other than 12 PCGs of mitochondria have been found in their respective genomes (Table 3).

In *W. bancrofti* and *B. malayi* alignment, 6 conserved regions were identified in which 5 regions were aligned to each other while only one was left. Maximum 33 ORFs were predicted in the 4500-11090 nucleotide region of *W. bancrofti* and a multiple C2 and transmembrane domain-containing protein 2 isoform X1 has been predicted again (also predicted in the alignment of *C. elegans* and *W. bancrofti*) in one of the ORF of 4500-11090 nucleotide region of *W. bancrofti*. 27 ORFs were predicted in 4458-11075 nucleotide region of *B. malayi*. In one region (1019-1041 nucleotide region of *W. bancrofti* and 11076-11098 nucleotide region of *B. malayi*) no ORF was predicted for both the genomes (Table 4). The neuronal IgCAM of *C. elegans* was also predicted in one of the ORF in the 11095-13644 nucleotide region of *W. bancrofti*. Two uncharacterized proteins namely BM_BM5154 and BM_BM126 and one hypothetical protein of *Wolbachia* endosymbiont of *Mansonella perstans* (A filarial nematode) were identified in the 1019-4456 nucleotide region of *B. malayi*.

Smart BLAST could not assign functions for all the predicted ORFs and only those which showed similarity with the known protein were listed (Table 2-4).

3.3 Identification of intergenic regions

In the alignment result of *C. elegans* and *W. bancrofti*, a total number of 20 conserved intergenic regions have been identified (Table 5) which have aligned with each other showing identity ranging between 68 to 74%. The identified intergenic regions which are part of the non-coding region have sequence length ranging between 104 bp (smallest) to 333 bp (largest). The alignment of *C. elegans* and *B. malayi* have identified 22 intergenic regions with identity ranging between 67.8 to 74 (Table 6). The identified conserved intergenic regions have sequence length ranging between 98 bp to 277 bp. In case of alignment of *W. bancrofti* and *B. malayi* only one conserved intergenic region has been identified and aligned with each other showing highest sequence identity of 87.5%. UCSC genome browser comparison of the three mitochondrial genomes also reveals the conservation among the different nematodal species. Regions of high conservation corresponds to the coding regions while intergenic regions have relatively less conservation (Figure 2).

3.4 ORF

A total number of 63, 58, and 40 ORFs were identified in the complete mt genome of *W. bancrofti* (Table S1), *B. malayi* (Table S2), and *C. elegans* (Table S3) respectively

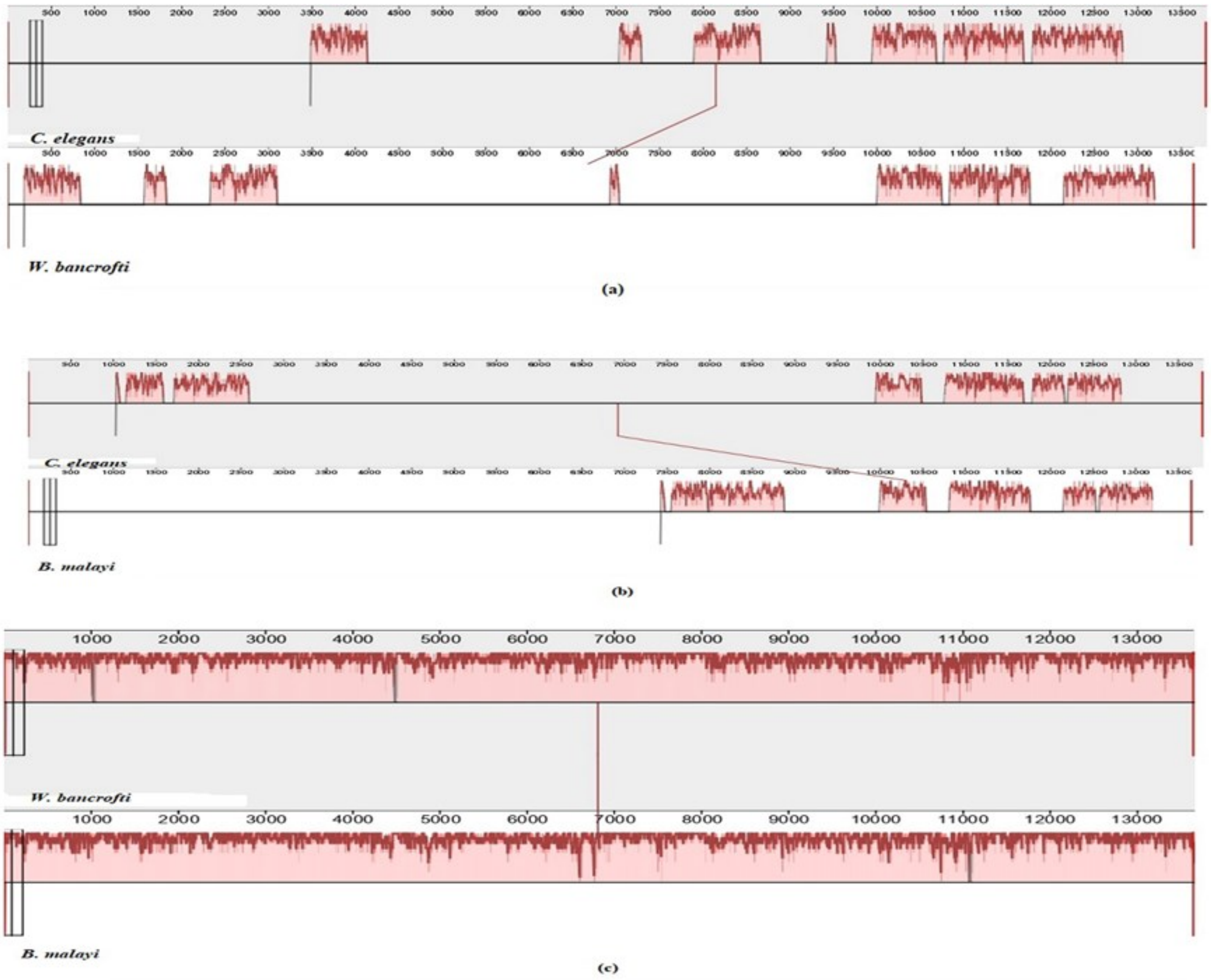


Figure 1 | (a) Alignment of *W. bancrofti* to *C. elegans*; (b): Alignment of *B. malayi* to *C. elegans*; (c): Alignment of *B. malayi* to *W. bancrofti* generated by the MAUVE viewer. The coloured bars inside the blocks are related to the level of sequence similarities. The line links the blocks with homology between two genomes that are aligned from top to bottom. The figure represents the genome alignment in which each panel represents one genome. The colourful blocks outline those regions of the genome sequence of one organism which have been aligned to the other genome sequence. The regions outside the coloured blocks/area of white regions depict the regions that have not been aligned, probably due to lineage specific sequences. The alignment panels of the *W. bancrofti* and *B. malayi* (c) genome alignment results are completely covered with the region of colour block depicting the highest level of sequence alignment as compared to the other two alignment results.

Table 1 | Comparison of nucleotide content of mtDNAs

Species	*W. bancrofti*	*B. malayi*	*C. elegans*
Family	*Onchocercinae*	*Onchocercinae*	*Rhabditidae*
Accession Number	NC_016186	AF538716	NC_001328
Length (bp)	13636	13657	13794
A %	20%	21%	31%
T%	55%	56%	47%
G%	18%	16%	14%
C%	7%	7%	8%
AT	75%	77%	78%

Table 2 | Alignment results of *W. bancrofti* to *C. elegans* (Backbone file results) showing the conserved regions generated by the Progressive Mauve program. ORFs predicted in the identified conserved regions followed by smart BLAST results for assigning the functions to ORFs. (NOTE- The differences in the no. of ORFs predicted and the results show that the smart BLAST could not predict functions of all the ORFs as no significant similarity was found with any known protein.)

S.No.	*C. elegans* nucleotide region (left end to right end)	No. of ORFs predicted	Smart BLAST results	*W. bancrofti* nucleotide region (left end to right end)	No. of ORFs predicted	Smart BLAST results
1.	3486-4136	2	Smart BLAST found no matches for both the ORFs	186-832	4	1.NADH dehydrogenase subunit 2 2. NADH dehydrogenase subunit 2
2.	7040-7286	1	Smart BLAST found no matches	1574-1819	1	hypothetical protein WUBG_19143 [*Wuchereria bancrofti*]
3.	7900-8654	4	1. cytochrome c oxidase subunit I 2. cytochrome c oxidase subunit I 3. cytochrome c oxidase subunit I 4. cytochrome c oxidase subunit I	2335-3094	4	1. Cytochrome oxidase subunit I 2. cbb3-type cytochrome c oxidase subunit I [*Wolbachia* endosymbiont of *Mansonella perstans*]
4.	9430-9521	0		6933-7028	0	
5.	9948-10677	2	Smart BLAST found no matches in both the ORFs	10007-10742	2	cytochrome c oxidase subunit 2
6.	10771-11682	2	Smart BLAST found no matches in both the ORFs	10837-11754	4	NADH dehydrogenase subunit 3
7.	11795-12823	0		12153-13187	6	1.NADH dehydrogenase subunit 5 2. NADH dehydrogenase subunit 5 3. neuRonal IGCAM [*Caenorhabditis elegans*]
8.	4137-7039	12	1.NADH dehydrogenase subunit 4 2. cytochrome b 3. cytochrome c oxidase subunit III 4. NADH dehydrogenase subunit 2 5. cytochrome c oxidase subunit III	0		
9.	7287-7899	2	1.NADH dehydrogenase subunit 4 2. NADH dehydrogenase subunit 4	0		
10.	8655-9429	4	1.cytochrome c oxidase subunit I 2. hypothetical protein DI535_26975]	0		
11.	9522-9947	0		0		
12.	10678-10770	0		0		
13.	11683-11794	0		0		
14.	0			833-1573	5	NADH dehydrogenase subunit 4
15.	0			1820-2334	2	NADH dehydrogenase subunit 4
	0			3095-6932	20	1.cbb3-type cytochrome c oxidase subunit I 2. cytochrome c oxidase subunit III 3. cytochrome b 4. cytochrome b 5. cytochrome c oxidase subunit 1 6. cytochrome b
16.	0			7029-10006	12	1.NADH dehydrogenase subunit 1 2. multiple C2 and transmembrane domain-containing protein 2 isoform X1
17.	0			10743-10836	0	
18.	0			11755-12152	0	

Table 3 | Alignment results of *B. malayi* to *C. elegans* (Backbone file results) showing the conserved regions generated by the Progressive Mauve program. ORFs predicted in the identified conserved regions followed by smart BLAST results for assigning the functions to ORFs. (NOTE- The differences in the no. of ORFs predicted and the smart BLAST results show that in the rest of the remaining ORFs no function could be assigned.)

S. No.	*C. elegans* nucleotide region (left end to right end)	No. of ORFs predicted	Smart BLAST results	*B. malayi* nucleotide region (left end to right end)	No. of ORFs predicted	Smart BLAST results
1.	1028-1069	0		7427-7466	0	
2.	1154-1583	1	Smart BLAST found no matches	7555-7976	1	Smart BLAST found no matches
3.	1715-2589	2	NADH dehydrogenase subunit 1	7997-8869	3	NADH dehydrogenase subunit 1
4.	9951-10485	1	Smart BLAST found no matches	9998-10534	3	1.cytochrome c oxidase subunit II 2. cytochrome oxidase subunit 2
5.	10763-11681	2	Smart BLAST found no matches in both the ORFs	10816-11758	4	NADH dehydrogenase subunit 3
6.	11795-12159	0		12158-12526	2	Smart BLAST found no matches in both the ORFs
7.	12210-12823	0		12576-13192	2	NADH dehydrogenase subunit 5
8.	1070-1153	0		0		
9.	1584-1714	0		0		
10.	2590-9950	25	1.NADH dehydrogenase subunit 4 2.NADH dehydrogenase subunit 2 3. NADH dehydrogenase subunit 4 4.cytochrome c oxidase subunit I 5.cytochrome c oxidase subunit I 6.ATP synthase F0 subunit 6 7. cytochrome c oxidase subunit I 8. cytochrome c oxidase subunit I 9.cytochrome b 10. cytochrome c oxidase subunit I 11. cytochrome c oxidase subunit III 12. cytochrome c oxidase subunit I 13. NADH dehydrogenase subunit 2 14. cytochrome c oxidase subunit III	0		
11.	10486-10762	1	Smart BLAST found no matches	0		
12.	11682-11794	0		0		
13.	12160-12209	0		0		
14.	0			7467-7554	0	
15.	0			7977-7996	0	
16.	0			8870-9997	5	No results found in any of the 5 ORFs
17.	0			10535-10815	0	
18.	0			11759-12157	0	
19.	0			12527-12575	0	

Table 4 | Alignment results of *B. malayi* to *W. bancrofti* (Backbone file results) showing the conserved regions generated by the Progressive Mauve program. ORFs predicted in the identified conserved regions followed by smart BLAST results for assigning the functions to ORFs. (NOTE- The differences in the no. of ORFs predicted and the smart BLAST results show that in the rest of the remaining ORFs no function could be assigned.)

S.No.	*W. bancrofti* nucleotide region (left end to right end)	No. of ORFs predicted	Smart BLAST results	*B. malayi* nucleotide region(left end to right end)	No. of ORFs predicted	Smart BLAST results
1	12-1018	4	1.NADH dehydrogenase subunit 2 2. NADH dehydrogenase subunit 2	3-1015	4	1.NADH dehydrogenase subunit 2 2. NADH dehydrogenase subunit 2
2	1042-4481	13	1.NADH dehydrogenase subunit 4 2. cytochrome oxidase subunit I 3. cbb3-type cytochrome c oxidase subunit I 4. cbb3-type cytochrome c oxidase subunit I [Wolbachia endosymbiont of Mansonella perstans] 5. cytochrome c oxidase subunit 1	1019-4456	17	1.Uncharacterized protein BM_BM5154 2. NADH dehydrogenase subunit 4 3. Uncharacterized protein BM_BM126 4. cytochrome c oxidase subunit I 5. cbb3-type cytochrome c oxidase subunit 6. hypothetical protein [*Wolbachia* endosymbiont of *Mansonella perstans*] 7. glutamate receptor ionotropic, kainate 3 8. cytochrome oxidase subunit 1
3	4500-11090	33	1.NADH dehydrogenase subunit 1 2. cytochrome c oxidase subunit 2 3. cytochrome c oxidase subunit III 4. cytochrome b 5. cytochrome b 6. cytochrome b 7. multiple C2 and transmembrane domain-containing protein 2 isoform X1	4458-11075	27	1.NADH dehydrogenase subunit 1 2. cytochrome c oxidase subunit III 3. cytochrome b 4. cytochrome c oxidase subunit III 5. cytochrome b 6. cytochrome c oxidase subunit II 7. cytochrome b
4	11095-13644	12	1.NADH dehydrogenase subunit 5 2.NADH dehydrogenase subunit 5 3.NADH dehydrogenase subunit 3 4.neuRonal IgCAM [*C. elegans*]	11099-13649	8	1. NADH dehydrogenase subunit 5 2. NADH dehydrogenase subunit 3
5	1019-1041	0		0		
6	0			11076-11098	0	

Table 5 | List of conserved intergenic regions identified by the alignment of *W. bancrofti* to *C. elegans* generated by Vista server

S. No.	*C. elegans* nucleotide region (left end CE right end)	*W. bancrofti* nucleotide region (left end right end)	Sequence length	Identity	
1.	1439-1588	7848-7997	152 bp	73.0%	Intergenic
2.	1785-1888	8080-8183	104 bp	71.2%	Intergenic
3.	1989-2130	8284-8425	142 bp	68.3%	Intergenic
4.	2286-2592	8583-8887	312 bp	67.9%	Intergenic
5.	4693-4818	4787-4912	126 bp	76.2%	Intergenic
6.	5233-5375	5327-5469	143 bp	70.6%	Intergenic
7.	6227-6344	6324-6441	118 bp	69.5%	Intergenic
8.	7016-7142	1550-1674	127 bp	74.0%	Intergenic
9.	7172-7295	1706-1833	128 bp	68.8%	Intergenic
10.	8004-8179	2445-2620	176 bp	73.3%	Intergenic
11.	8340-8672	2781-3113	333 bp	70.3%	Intergenic
12.	8703-8863	3144-3307	165 bp	70.9%	Intergenic
13.	8893-9074	3334-3516	183 bp	70.5%	Intergenic
14.	9109-9208	3550-3649	100 bp	70.0%	Intergenic
15.	9927-10148	9984-10207	224 bp	69.2%	Intergenic
16.	10183-10370	10242-10426	188 bp	68.6%	Intergenic
17.	10901-11029	10972-11095	132 bp	68.9%	Intergenic
18.	11221-11328	11283-11389	109 bp	70.6%	Intergenic
19.	11979-12135	12340-12500	161 bp	70.2%	Intergenic
20.	12493- 12614	12854- 12975	125 bp	68.0%	Intergenic

Table 6 | List of conserved intergenic regions identified by the alignment of *B. malayi* to *C. elegans* generated by Vista server

S. No.	CE nucleotide region (left end to right end)	BM nucleotide region (left end to right end)	Sequence Length	Identity	
1.	1161-1304	7566-7699	144 bp	68.1%	Intergenic
2.	1434-1588	7829-7982	156 bp	72.4%	Intergenic
3.	2002-2118	8283-8399	117 bp	69.2%	Intergenic
4.	2283-2461	8564-8742	179 bp	68.2%	Intergenic
5.	4562-4661	4618-4717	100 bp	72.0%	Intergenic
6.	4687-4807	4737-4857	121 bp	71.9%	Intergenic
7.	4831-4988	4881-5038	158 bp	67.7%	Intergenic
8.	5080-5177	5130-5227	98 bp	70.4%	Intergenic
9.	5221-5382	5271-5429	162 bp	68.5%	Intergenic
10.	6946-7135	1457-1646	190 bp	73.2%	Intergenic
11.	7167-7302	1678-1820	143 bp	69.9%	Intergenic
12.	8004-8179	2422-2597	176 bp	71.6%	Intergenic
13.	8442-8656	2860-3074	221 bp	70.6%	Intergenic
14.	8701-8852	3119-3270	152 bp	72.4%	Intergenic
15.	8928-9034	3346-3452	107 bp	72.0%	Intergenic
16.	9098-9213	3516-3631	116 bp	69.0%	Intergenic
17.	9941-10114	9988-10161	174 bp	72.4%	Intergenic
18.	10186-10342	10233-10387	159 bp	67.3%	Intergenic
19.	10739-11012	10792-11067	277 bp	74.0%	Intergenic
20.	11180-11328	11241-11394	154 bp	70.1%	Intergenic
21.	11970-12128	12336-12494	159 bp	71.1%	Intergenic
22.	12202-12344	12568-12707	143 bp	67.8%	Intergenic

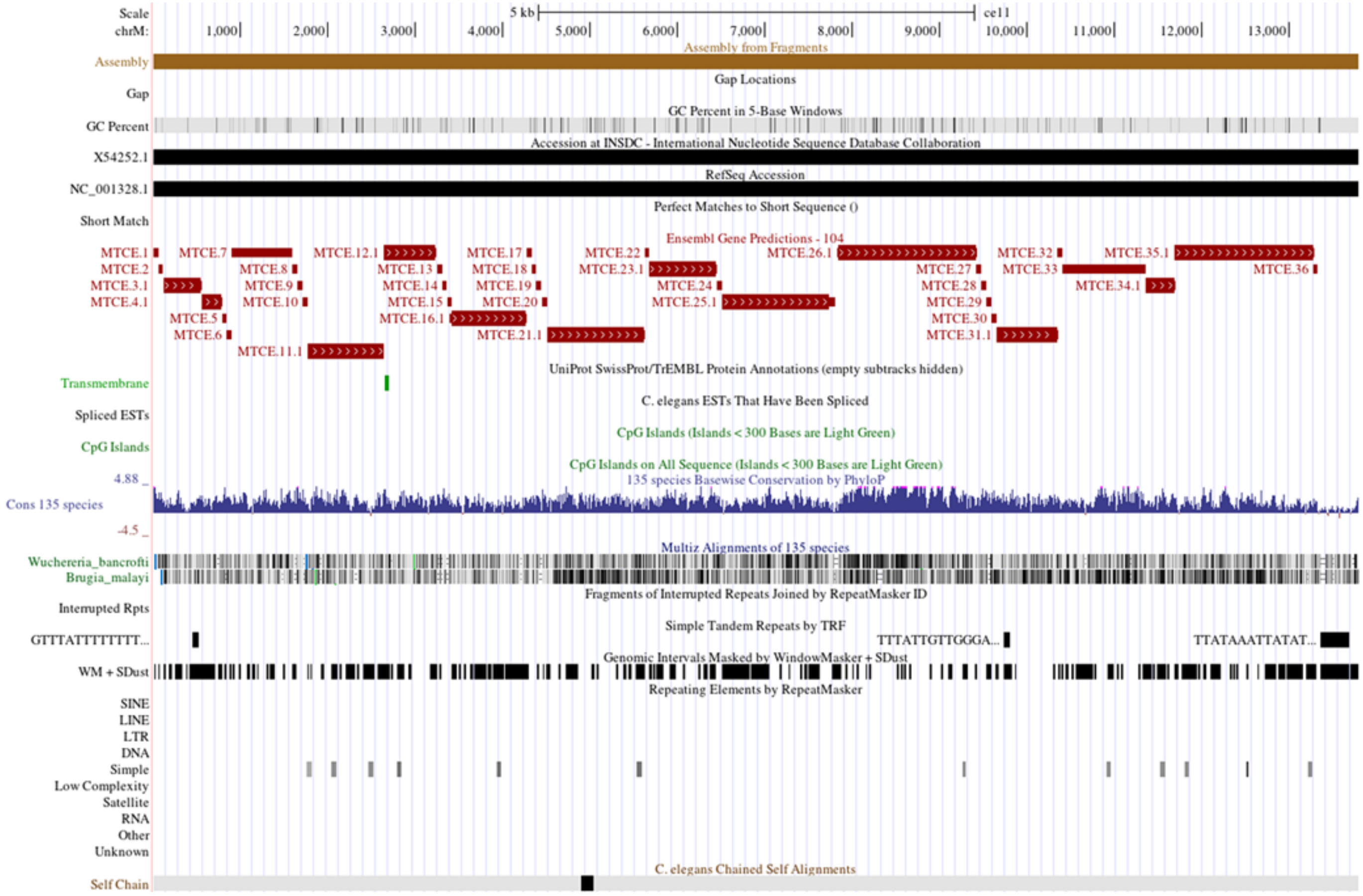


Figure 2 | The results depict the alignment of mitochondrial genomic data of *W. bancrofti* and *B. malayi* among the 135 available species for alignment in the browser. The peaks represent the degree of conservation among the two aligned genomes. The high-conserved areas are represented by dark regions in the genome panels of the species. Image prepared by UCSC genome browser (http://genome.ucsc.edu).

(Figure 3). After subjecting each ORF to BLASTP (protein-protein BLAST) only 28, 33, and 26 ORFs in *W. bancrofti*, *B. malayi* and *C. elegans* respectively had the functions while the remaining ORFs displayed no sequence similarity to any known proteins. Out of 28 ORFs in *W. bancrofti*, 17 ORFs were of already known 12 PCGs of mitochondria. While the remaining 11 ORFs were newly reported (Table 7). Similarly, in *B. malayi*, 17 ORFs showed the conserved genes present in mitochondria while the remaining 16 showed new results (Table 8). In the case of *C. elegans*, 9 results were new while the rest 17 ORFs showed similarity with the already known conserved genes of *C. elegans* (Table 9) (Figure 4). Although the ORF finder listed some of these regions as ORFs, they are most probably the non-coding regions of the mitochondrial genome. The ORFs and their subsequent blast were done to ascertain whether there is any difference at the mitochondrial gene coding level in these organisms.

3.5 Multiple sequence alignment and Phylogenetic relationships

Multiple sequence alignment results displayed by CLUSTALW for alignment of mt genomes of *C. elegans* with *W. bancrofti* and *B. malayi* shows significantly less alignment score (42) as compared to the alignment score of *W. bancrofti* and *B. malayi* (88).

The phylogenetic tree was constructed using the complete mt genome sequences of the species. The maximum likelihood method (Figure 5) clustered *W. bancrofti* and *B. malayi* in one distinct phylogenetic group while *C. elegans* was present in a different group. While these organisms are grouped into different clades than the *C. elegans*, they seem to share a common ancestor. These results showed that the two parasitic nematodes have more evolutionary relatedness to each other than *C. elegans*. All these three nematodes may have evolved from a common ancestor which explains the similarity in the mitochondrial genome of all three species.

3.6 Comparison of protein coding genes

All three nematode species contain the same 12 PCGs. As reported earlier, all the species lack ATP synthase subunit 8. For the validation of the amino acid sequence reterived from NCBI and PDB database alignment performed by pBLAST and 100% sequence similarity have been obtained for all the 12 PCGs. The alignment of mitochondrial protein coding genes of all three species shows the higher similarity between *W. bancrofti* and *B. malayi*. The sequence level similarity between genes from *C. elegans* and the other two organisms is less owing to the specific deletions and insertions in each

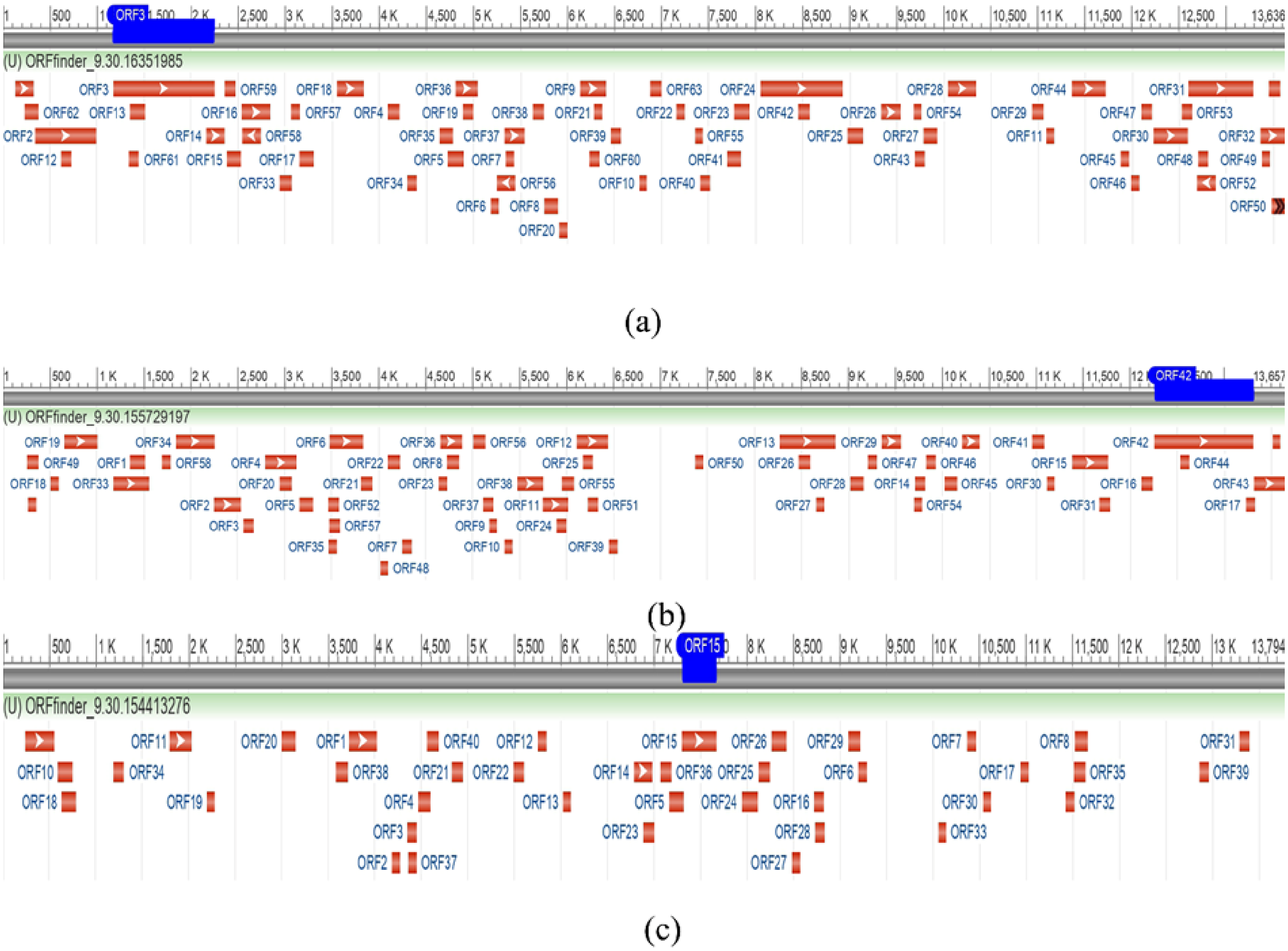


Figure 3 | Pictorial representation of all the ORFs predicted by ORF Finder tool of NCBI in the complete mitochondrial genome of (a) *W. bancrofti*; (b) *B. malayi*; (c) *C. elegans*.

Table 7 | List of new ORFs in *W. bancrofti* (Protein BLAST results)

S.NO.	Length of ORF (Nucleotide)	Length of ORF (Amino acid)	Function
1.	93	30	Helix-loop-helix protein 3
2.	87	28	Large envelope protein
3.	84	27	Vesicle transport protein GOT1A
4.	84	27	Meiotically up-regulated gene 73 protein
5.	81	26	Endoplasmic reticulum-Golgi intermediate compartment protein 1
6.	81	26	Cytotoxic and regulatory T-cell molecule
7.	81	26	Cytochrome P450 85A1
8.	78	25	Protein Opaque1
9.	78	25	HAUS augmin-like complex subunit 8
10.	78	25	Olfactory receptor
11.	78	25	V-type proton ATPase 116 kDa subunit a2

Table 8 | List of new ORFs in *B. malayi* (Protein BLAST results)

S.No.	Length of ORF (Nucleotide)	Length of ORF (Amino acid)	Function assignment
1.	93	30	UPF0295 protein GK0479
2.	93	30	Metabotropic glutamate receptor-like protein A
3.	93	30	Neuromedin-B receptor
4.	90	29	Guanylate cyclase beta
5.	87	28	Protoheme IX farnesyltransferase
6.	84	27	Serine/threonine-protein kinase ppk1
7.	81	26	Schlafen family member 12-like
8.	81	26	AB hydrolase superfamily protein B1A11.02
9.	81	26	Serine/threonine-protein kinase irlC
10.	81	26	Epidermal patterning factor-like protein 6
11.	81	26	Endoplasmic reticulum junction formation protein lunapark
12.	78	25	Pre-rRNA-processing protein IPI1
13.	78	25	DNA-directed RNA polymerase subunit beta
14.	78	25	Cell division protein FtsQ
15.	78	25	Dihydroneopterin monophosphate aldolase
16.	78	25	C-C chemokine receptor type 2

Table 9 | List of new ORFs in *C. elegans* (Protein Blast Results)

S.No	Length of ORF (Nucleotide)	Length of ORF (Amino acid)	Function
1.	93	30	Elongation factor G
2.	93	30	Phosphoglucosamine mutase
3.	93	30	Putative phosphoesterase GWCH70_0799
4.	90	29	Envelope glycoprotein O
5.	87	28	Delta (7)-sterol 5(6)-desaturase
6.	84	27	Mitochondrial aspartate-glutamate transporter AGC1
7.	81	26	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
8.	81	26	Protein glycosylation K
9.	78	25	Uncharacterized 24 kDa protein

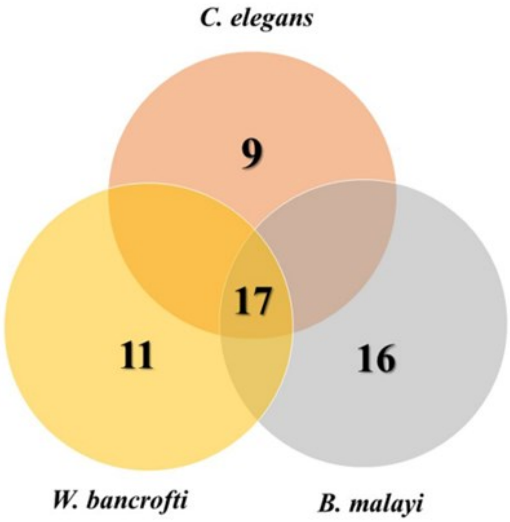


Figure 4 | Venn diagram representing numbers of functional common as well as unique ORFs in complete mt genomes of *C. elegans*, *W. bancrofti*, and *B. malayi*.

gene supporting this observation (Table S4). The sequence comparison shows the highest diversity of gene sequences among cytochrome C oxidase subunit 3 (*C. elegans* vs *W. bancrofti*) and NADH dehydrogenase subunit 6 (*C. elegans* vs *B. malayi*). Codon usage analysis using the sequence manipulation suite [17] for a single gene (NADH dehydrogenase subunit 1) reveals some important differences in codon usage in these three organisms. For proline, *C. elegans* preferred the codon CCA while the other two preferred CCT. Similarly, in *C. elegans*, CAA is preferred for glutamine, but *W. bancrofti* and *B. malayi* uses both CAG and CAA. For threonine, both *W. bancrofti* and *B. malayi* uses ACT while *C. elegans* prefer ACA and ACT (Table 10). These codon biases suggest that the parasitic nature of the *W. bancrofti* and *B. malayi* have been adapted for different codons for some of the amino acids.

4. Discussion

In this study, an attempt has been made to compare the mt genomes of filarial parasites *Wuchereria bancrofti* and *Brugia malayi* to a free-living nematode *C. elegans*. Access to the publicly available multiple genomes in different databases has offered improvised gene prediction by comparing the conserved genomic regions of multiple species [18].

The AT nucleotide content of reference genome *C. elegans* is 78 %. The AT nucleotide content of *W. bancrofti* was calculated to be 75 % consistent with the earlier reports [19]. While for *B. malayi*, it is 77% slightly more than the previous report of 75.5 % [20].

With the help of complete mt genome alignment, a total of 13 conserved regions have been identified in all three species. These regions are having variable number of ORFs within them. Meanwhile, a total of 20 and 22 conserved intergenic regions in *W. bancrofti* and *B. malayi* respectively have also been identified which are part of the non coding regions that can be found within genes. While in the complete mt genome a total number of 32,35 and 26 functional ORF have been predicted in *W. bancrofti*, *B. malayi* and *C. elegans* respectively. The total number of ORFs of these species has not been reported earlier in any previous study. Among all the identified ORFs, 17 ORFs were found to be common in all the three genomes studied.

The MSA is a fundamental part of comparative sequence analysis [21] that has been performed with the help of CLUSTALW in this study. The difference in the CLUSTALW alignment score depicts the sequence identity difference between *C. elegans* and the other two species. The sequence similarity depicted by the alignment score of *C. elegans* to *W. bancrofti* and *B. malayi* was almost half of the alignment score of *W. bancrofti* and *B. malayi* indicating that their sequence similarity/relatedness is much higher than each other as compared to *C. elegans*.

The complete mt genome of total of 62 species was used to construct the phylogenetic tree through the Maximum likelihood method. All the species compared were clustered according to their free-living nature and host preferences (either being a plant, animal, or human parasite). The phylogenetic analysis depicts a close relationship between *W. bancrofti* and *B. malayi*. Both the species were present in the same phylogenetic clade and causes the same disease (LF) in humans. Such a similar finding regarding *W. bancrofti* and *B. malayi* has also been reported previously [11] [19] by using the concatenated nucleotide sequences of only twelve PCGs of nematodes. *C. elegans* is grouped into separate clade but share a common ancestor with the other two studied nematodes.

In a study of the mt genome sequence of 5 nematode species that were either *Wolbachia* dependent or independent parasites, it was observed that the species were remarkably similar on their sequence level despite their dependency on *Wolbachia* for their survival [11]. While the mitochondrial sequences of *W. bancrofti* and *B. malayi* are found to be similar in respect of their size, AT content and encode the same 12 PCGs, we have found the unique conserved non-coding regions that are present only in either of the genome. These unique regions may carry the difference between the free living and the parastic organism.

5. Concluding Remarks

W. bancrofti and *B. malayi* are genetically similar to each other but are divergent from the model species *C. elegans* even though they have evolved from a common ancestor. There was an early divergence of these two species from *C. elegans* but still, they all contain the same 12 PCGs in their mt genome. Analysis of the mt genome of *W. bancrofti* and *B. malayi* showed that it contains some unique conserved regions. These conserved regions may not contain any protein coding genes but are conserved at their sequence level. Some new unique (partial) ORFs have also been

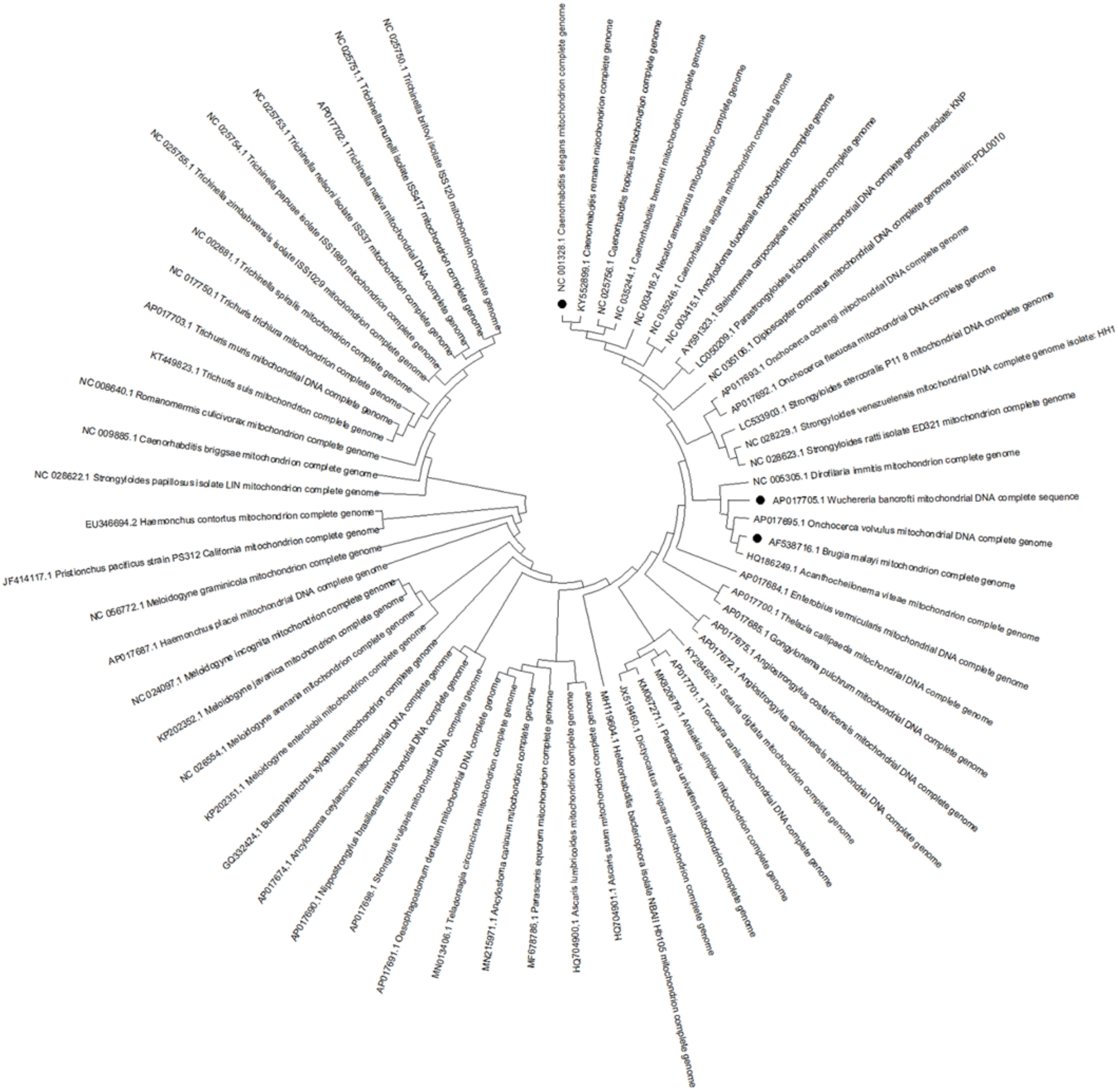


Figure 5 | Evolutionary analysis by Maximum Likelihood method in MEGA X. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model. All the species compared were clustered according to their free-living nature and host preferences (either being a plant, animal, or human parasite). The tree constructed depicts a close relationship between *W. bancrofti* and *B. malayi*. Both species were present in the same phylogenetic clade. *C. elegans* is grouped into separate clade but share a common ancestor with the other two studied nematodes.

Table 10 | Codon usage analysis using the sequence manipulation suite [17] for a single gene (NADH dehydrogenase subunit 1)

S.No.	Amino Acid	Codons	Fraction		
			CE	WB	BM
1.	Ala	GCG	0.08	0.11	0.20
		GCA	0.23	0.0	0.0
		GCT	0.54	0.89	0.50
		GCC	0.15	0.00	0.30
2.	Cys	TCT	1.0	0.93	0.83
		TGC	0.0	0.07	0.17
3.	Asp	GAT	1.0	1.0	1.0
		GAC	0.00	0.0	0.0
4.	Glu	GAG	0.33	0.67	0.25
		GAA	0.67	0.33	0.75
5.	Phe	TTT	1.0	0.98	0.98
		TTC	0.0	0.02	0.02
6.	Gly	GGG	0.25	0.05	0.05
		GGA	0.50	0.81	0.05
		GGT	0.25	0.10	0.76
		GGC	0.00	0.05	0.14
7.	His	CAT	1.00	0.5	1.0
		CAC	0.00	0.5	0.0
8.	Ile	ATA	0.26	0.19	0.15
		ATT	0.62	0.77	0.85
		ATC	0.12	0.04	0.00
9.	Lys	AAG	0.11	0.83	0.5
		AAA	0.89	0.17	0.5
10.	Leu	TTG	0.12	0.49	0.34
		TTA	0.67	0.40	0.45
		CTG	0.00	0.00	0.02
		CTA	0.05	0.00	0.07
		CTT	0.16	0.11	0.11
		CTC	0.00	0.00	0.00
11.	Met	ATG	1.00	1.00	1.00
12.	Asn	AAT	0.88	1.0	1.0
		AAC	0.13	0.0	0.0
13.	Pro	CCG	0.00	0.00	0.00
		CCA	0.63	0.14	0.00
		CCT	0.13	0.86	0.86
		CCC	0.25	0.00	0.14
14.	Gln	CAG	0.00	0.40	0.60
		CAA	1.00	0.60	0.40
15.	Arg	AGG	0.14	0.15	0.00
		AGA	0.52	0.31	0.30
		CGG	0.05	0.08	0.10
		CGA	0.00	0.00	0.00
		CGT	0.29	0.46	0.60
		CGC	0.00	0.00	0.00
16.	Ser	AGT	0.18	0.25	0.35
		AGC	0.05	0.00	0.00
		TCG	0.00	0.00	0.00
		TCA	0.32	0.08	0.08
		TCT	0.41	0.63	0.58
		TCC	0.05	0.04	0.00
17.	Thr	ACG	0.14	0.00	0.00
		ACA	0.43	0.00	0.00
		ACT	0.29	1.00	1.00
		ACC	0.14	0.00	0.00
18.	Val	GTG	0.05	0.18	0.00
		GTA	0.32	0.09	0.00
		GTT	0.64	0.73	0.94
		GTC	0.00	0.001	0.60
19.	Trp	TGG	0.00	1.0	1.0
20.	Tyr	TAT	0.81	0.89	0.83
		TAC	0.19	0.11	0.17
21.	END	TGA	0.67	0.00	1.00
		TAG	0.00	0.00	0.00
		TAA	0.33	1.00	0.00

identified in these species which have not been reported earlier. The analysis of the intergenic regions between these organisms showed a range of conserved regions between all three nematodes. The study identified these common conserved regions in these mitochondrial genomes which may explain their evolutionary and lifestyle differences.

Supplementary material:

Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Table S4

Acknowledgements

SJ, LK and KR thank University Grant Commission (UGC) for providing the financial support in the form of fellowship. We are also thankful to Dr. Shailesh Kumar (Staff Scientist III, NIPGR) for their kind suggestions for this work.

References

[1] X. Qing, K. Kulkeaw, S. Wongkamchai, S.K.-W. Tsui, Front. Ecol. Evol. 9 (2021) 25. DOI: 10.3389/fevo.2021.637805

[2] A. Coghlan, R. Tyagi, J.A. Cotton, N. Holroyd, B.A. Rosa, I.J. Tsai, D.R. Laetsch, R.N. Beech, T.A. Day, K. Hallsworth-Pepin, H.-M. Ke, T.-H. Kuo, T.J. Lee, J. Martin, R.M. Maizels, P. Mutowo, P. Ozersky, J. Parkinson, A.J. Reid, N.D. Rawlings, D.M. Ribeiro, L.S. Swapna, E. Stanley, D.W. Taylor, N.J. Wheeler, M. Zamanian, X. Zhang, F. Allan, J.E. Allen, K. Asano, S.A. Babayan, G. Bah, H. Beasley, H.M. Bennett, S.A. Bisset, E. Castillo, J. Cook, P.J. Cooper, T. Cruz-Bustos, C. Cuéllar, E. Devaney, S.R. Doyle, M.L. Eberhard, A. Emery, K.S. Eom, J.S. Gilleard, D. Gordon, Y. Harcus, B. Harsha, J.M. Hawdon, D.E. Hill, J. Hodgkinson, P. Horák, K.L. Howe, T. Huckvale, M. Kalbe, G. Kaur, T. Kikuchi, G. Koutsovoulos, S. Kumar, A.R. Leach, J. Lomax, B. Makepeace, J.B. Matthews, A. Muro, N.M. O'Boyle, P.D. Olson, A. Osuna, F. Partono, K. Pfarr, G. Rinaldi, P. Foronda, D. Rollinson, M.G. Samblas, H. Sato, M. Schnyder, T. Scholz, M. Shafie, V.N. Tanya, R. Toledo, A. Tracey, J.F. Urban, L.-C. Wang, D. Zarlenga, M.L. Blaxter, M. Mitreva, M. Berriman, International Helminth Genomes Consortium, Nat. Genet. 51 (2019) 163–174. DOI: 10.1038/s41588-018-0262-1

[3] A.L. Scott, E. Ghedin, Parasitol. Int. 58 (2009) 6–11. DOI: 10.1016/j.parint.2008.09.003

[4] E.M.A. Kern, T. Kim, J.-K. Park, Front. Ecol. Evol. 8 (2020) 250. DOI: 10.3389/fevo.2020.00250

[5] J.L. Boore, Nucleic Acids Res. 27 (1999) 1767–1780. DOI: 10.1093/nar/27.8.1767

[6] M. Hu, R.B. Gasser, Trends Parasitol. 22 (2006) 78–84. DOI: 10.1016/j.pt.2005.12.003

[7] J.-F. Gao, Q. Zhao, G.-H. Liu, Y. Zhang, Y. Zhang, W.-T. Wang, Q.-C. Chang, C.-R. Wang, X.-Q. Zhu, Gene 541 (2014) 92–100. DOI: 10.1016/j.gene.2014.03.017

[8] L. Wei, Y. Liu, I. Dubchak, J. Shon, J. Park, J. Biomed. Inform. 35 (2002) 142–150. DOI: 10.1016/S1532-0464(02) 00506-3

[9] R.C. Hardison, PLoS Biol. 1 (2003) e58. DOI: 10.1371/journal.pbio.0000058

[10] A.G. Hise, I. Gillette-Ferguson, E. Pearlman, Cell. Microbiol. 6 (2004) 97–104. DOI: 10.1046/j.1462-5822.2003.00350.x

[11] S.N. McNulty, A.S. Mullin, J.A. Vaughan, V.V. Tkach, G.J. Weil, P.U. Fischer, BMC Genomics 13 (2012) 145. DOI: 10.1186/1471-2164-13-145

[12] A.C.E. Darling, B. Mau, F.R. Blattner, N.T. Perna, Genome Res. 14 (2004) 1394–1403. DOI: 10.1101/gr.2289704

[13] K.A. Frazer, L. Pachter, A. Poliakov, E.M. Rubin, I. Dubchak, Nucleic Acids Res. 32 (2004) W273–W279. DOI: 10.1093/nar/gkh458

[14] W.J. Kent, C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler, Genome Res. 12 (2002) 996–1006. DOI: 10.1101/gr.229102

[15] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673–4680. DOI: 10.1093/nar/22.22.4673

[16] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, Mol. Biol. Evol. 35 (2018) 1547–1549. DOI: 10.1093/molbev/msy096

[17] P. Stothard, BioTechniques 28 (2000) 1102, 1104. DOI: 10.2144/00286ir01

[18] W. Miller, K.D. Makova, A. Nekrutenko, R.C. Hardison, Annu. Rev. Genomics Hum. Genet. 5 (2004) 15–56. DOI: 10.1146/annurev.genom.5.061903.180057

[19] A. Ramesh, S.T. Small, Z.A. Kloos, J.W. Kazura, T.B. Nutman, D. Serre, P.A. Zimmerman, Mol. Biochem. Parasitol. 183 (2012) 32–41. DOI: 10.1016/j.molbiopara.2012.01.004

[20] E. Ghedin, S. Wang, D. Spiro, E. Caler, Q. Zhao, J. Crabtree, J.E. Allen, A.L. Delcher, D.B. Guiliano, D. Miranda-Saavedra, S.V. Angiuoli, T. Creasy, P. Amedeo, B. Haas, N.M. El-Sayed, J.R. Wortman, T. Feldblyum, L. Tallon, M. Schatz, M. Shumway, H. Koo, S.L. Salzberg, S. Schobel, M. Pertea, M. Pop, O. White, G.J. Barton, C.K.S. Carlow, M.J. Crawford, J. Daub, M.W. Dimmic, C.F. Estes, J.M. Foster, M. Ganatra, W.F. Gregory, N.M. Johnson, J. Jin, R. Komuniecki, I. Korf, S. Kumar, S. Laney, B.-W. Li, W. Li, T.H. Lindblom, S. Lustigman, D. Ma, C.V. Maina, D.M.A. Martin, J.P. McCarter, L. McReynolds, M. Mitreva, T.B. Nutman, J. Parkinson, J.M. Peregrin-Alvarez, C. Poole, Q. Ren, L. Saunders, A.E. Sluder, K. Smith, M. Stanke, T.R. Unnasch, J. Ware, A.D. Wei, G. Weil, D.J. Williams, Y. Zhang, S.A. Williams, C. Fraser-Liggett, B. Slatko, M.L. Blaxter, A.L. Scott, Science 317 (2007) 1756–1760. DOI: 10.1126/science.1145406

[21] O. Penn, E. Privman, G. Landan, D. Graur, T. Pupko, Mol. Biol. Evol. 27 (2010) 1759–1767. DOI: 10.1093/molbev/msq066